Master Thesis

TVVR 24/5002

Assessing Soil Erosion in the Nyagako Sub-catchment, Rwanda

Integrating GIS Analysis, Turbidity Measurements, and Sediment Fingerprinting

---

Philip Hultemar

Division of Water Resources Engineering
Department of Building and Environmental Technology
Lund University

Assessing Soil Erosion in the Nyagako Sub-catchment, Rwanda

Integrating GIS Analysis, Turbidity Measurements, and Sediment Fingerprinting

By:

Philip Hultemar

Master Thesis

Division of Water Resources Engineering
Department of Building & Environmental Technology
Lund University
Box 118
221 00 Lund, Sweden

Cover photo by Author

Water Resources Engineering
TVVR-24/5002
ISSN 1101-9824

Lund 2024
www.tvrl.lth.se

Master Thesis
Division of Water Resources Engineering
Department of Building & Environmental Technology
Lund University

English title:	Assessing Soil Erosion in the Nyagako Sub-catchment, Rwanda Integrating GIS Analysis, Turbidity Measurements, and Sediment Fingerprinting
Author:	Philip Hultemar
Supervisor:	Dr. Erik Nilsson
Assistant Supervisors:	Prof. Magnus Larson Mr. Leonard Nzabonantuma
Examiner:	Associate Prof. Linus Zhang
Language	English
Year:	2024
Keywords:	Soil erosion; Sediment fingerprinting analysis; GIS; Rwanda; Turbidity; MixSIAR; Deconvoluted; RUSLE

Acknowledgments

I would like to express my gratitude for the support and assistance that made this project possible. I am deeply thankful to the Crafoord Foundation for the travel grant and to Lund University for their support throughout the process of this master thesis.

Thank you Dr. Maarten Wynants, for your guidance and expertise in the sediment fingerprinting analysis, this project would not have been possible without your support.

Prof. Umaru Garba Wali, thank you for welcoming me to the University of Rwanda and providing all the necessary equipment and laboratory space. I am also grateful to the technical staff at the university; Mr. Birori Mudakikwa Mardochée for assisting with the turbidity measurements and providing valuable insights, Mr. Louis Theo Dushimirimana in the soil laboratory for facilitating the sample preparation, Mr. Samson Ntirenganyafor assisting with the XRF-analysis and to Mr. Serge Mugabo for translating during fieldwork. Many thanks to Mr. Ibuka Joseph for your local guidance and ensuring my safety during fieldwork.

A special thanks to Dr. Erik Nilsson for being an exceptional supervisor, introducing me to Rwanda and the SIDA funded capacity building program “Enhancing Water Resources and Environmental Management for Sustainable Livelihoods in Rwanda” (E-WREM-SLR) between Lund University and the University of Rwanda, and providing invaluable discussions and feedback. To my assistant supervisors; Prof. Magnus Larson, thank you for your wise inputs and discussions, and Mr. Leonard Nzabonantuma, I am grateful for your support in Rwanda and for welcoming me into your family.

Lastly, many thanks to my family and Vera for all the support and shared memories.

Abstract

Soil erosion is a significant challenge in Rwanda, particularly within the Nyagako sub-catchment, where this study focuses on a delimited area. This research investigates the sources and contributing factors of soil erosion using GIS analysis, turbidity measurements, and sediment fingerprinting with a deconvoluted MixSIAR (D-MixSIAR) model in R. The objectives were to identify erosion-prone areas, assess various factors' impacts, and evaluate the methodology's efficacy. Results revealed that the most downstream region (Zone C) is the most erosion-prone, contributing 70-80% of the total sediment, while CA (alluvium and colluvium deposits) as the most erosion-prone source per hectare across the zones. The most upstream region (Zone A) was the second-highest contributor, with GG as the major contributing source, accounting for 60.9% at Zone A's outlet and 11.3% at the final outlet. A strong positive correlation was found between soil erosion and the cultivation-forest ratio, while steep gradients showed a negative correlation. Turbidity analysis highlighted the impact of mining activities and rainfall events on erosion rates, with backtracking indicating significant contributions from sub-catchments A1 and A5. This research offers a methodological approach to assess soil erosion within challenging environments with limited resources, providing valuable insights for future soil erosion assessment.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CA	Colluvium and Alluvium deposits
GG	Acidic rocks (Granite, Gneiss, Tonalite)
QS	Quartzite and Shale formations (Schists, Argillites, Ardoises, Phyllites)
NTU	Nephelometric Turbidity Units
TSS	Total Suspended Solids
XRF	X-ray Fluorescence
RUSLE	Revised Universal Soil Loss Equation
MCMC	Markov Chain Monte Carlo
JAGS	Just Another Gibbs Sampler
GIS	Geographic Information System
D-MixSIAR	Deconvolutional MixSIAR
LULC	Land Use and Land Cover
NNYU	Nile Nyabarongo Upper Catchment
GLOWS-FIU	Global Water for Sustainability Program - Florida International University
R	Statistical computing language and environment
SD	Standard Deviation
PCA	Principal Component Analysis
DF	Degrees of Freedom
SC	Soil Campaign
FAO	Food and Agriculture Organization
SECM	Soil Erosion Control Measures

1. Introduction

1.1. Soil Erosion: Global Issue and Local Challenges in Rwanda

Soil erosion is a severe global challenge, affecting millions of hectares and leading to considerable loss in agricultural productivity, soil carbon reserves, and increasing sedimentation in aquatic ecosystems. A particularly alarming issue is the disparity between the rates of soil erosion and soil formation, endangering soil health and future sustainability on a global scale. Factors such as intensive farming, deforestation, and changes in land use significantly contribute to this problem by removing the natural vegetation that safeguards against erosion (Food and Agriculture Organization of the United Nations & Intergovernmental Technical Panel on Soils, 2015). As climate change amplifies extreme weather events, an increase in soil erosion globally is expected, exacerbating its impact on ecosystem services and human well-being. Additionally, this intensification of soil erosion is further compounded by land use changes, which also contribute to the alteration of biogeochemical cycles, thus feeding back into the cycle of climate change (Eekhout & de Vente, 2022).

In Rwanda, soil erosion is considered as the most serious environmental issue (Government of Rwanda & IUCN, 2022), with numerous regions in Rwanda experiencing soil erosion rate at critical levels, where the national averages reach 25 tonnes per hectare each year. Soil erosion causes considerable losses in agricultural productivity and topsoil, leading to a significant economic impact on the agricultural GDP and necessitating costly soil fertility replenishments through fertilizers. This environmental challenge not only threatens the nation's food security but also imposes a substantial financial burden on restoring soil health and productivity (Government of Rwanda & IUCN, 2022). Rwanda, situated in Central Africa, features a varied terrain from the Congo-Nile ridge's steep slopes to the less rugged eastern plains. The country's geomorphology includes high-altitude areas like the volcanic Birunga chain, with the highest peak at Karisimbi volcano (4,507 meters), contributing to its steep river gradients and pronounced topographical diversity. Rwanda's landscape is shaped

by differential erosion, with "hard" quartzitic rocks forming steep, prominent mountains above softer, schist-based valleys. The climate is temperate featuring two main rainy seasons and the potential for varied precipitation patterns across different regions due to its altitude variation. This intricate combination of steep topography, diverse geology, and distinct climatic zones makes Rwanda highly susceptible to soil erosion, posing challenges for land management and conservation. Only 26% of the land at risk of erosion is currently protected through measures like contour bank terraces and forest coverage, leaving the majority of vulnerable land still exposed to erosion threats (Government of Rwanda & IUCN, 2022).

Heavy rainfall, saturated local soils, followed by a small earthquake is enough to trigger a catastrophic debris of avalanches, torrents, and earthflows within the following 24 hours. An example of this kind of catastrophic event occurred in western Rwanda in early May 1988, leading to significant loss of life, property, and livelihood. exacerbated soil loss, averaging 34 tons per hectare on cropped runoff plots (Byers III, 1992). The impacted area's Nyamutera River transported 567,000 tons of suspended sediment to its mouth between May 7 and May 13, indicating a significant basin-wide decrease in sediment yield. The susceptibility of regions like Nyakinama to similar extreme weather events, exacerbating existing soil loss issues. Given that contemporary land use practices worsened the severity of the 1988 event. There is an urgent need for applied research to develop technologies that reduce soil loss, improve soil fertility, and mitigate the impacts of heavy rainfall events.

Implementing sustainable management and agricultural practices in Rwanda, especially in areas like the Sebeya catchment area, presents significant challenges. Farmers in this region face severe soil erosion consequences, including property damage, loss of life, and reduced crop yields, yet the broad adoption of soil conservation measures remains limited (Majoro et al., 2020). Barriers related to finances and knowledge hinder the widespread application of soil erosion control measures (SECM), emphasizing the necessity for government intervention. Such intervention should aim to equip farmers with both the technical know-how and the financial support needed to adopt and implement key practices that

promote sustainable agriculture effectively (Majoro et al., 2023). Effective soil erosion management requires both preventive legislation for afforestation and a shift towards dry season farming to protect vulnerable soils. Additionally, empowering affected communities with the knowledge and financial support to implement early-stage erosion control measures is crucial for long-term environmental sustainability (Igwe et al., 2017).

A previous study in 2016 investigated factors influencing the adoption of soil conservation techniques in the Gatebe sector of northern Rwanda, where poor land management practices have led to soil erosion and declining soil fertility. Through surveys and data analysis, it was found that while many farmers were aware of the causes of soil erosion, challenges such as over-cultivation on steep farmlands, limited resources, and insufficient training hindered effective soil conservation. Land acquisition methods significantly influenced adoption rates, with those owning or inheriting land being more likely to adopt conservation practices. The study underscores the importance of policymakers enhancing farmer awareness and promoting efficient soil conservation strategies through training and support programs (Nahayo, Pan, & Joseph, 2016). However, there is a high commitment to addressing soil erosion in Rwanda. Strategic planning, targeted interventions, and investment in sustainable land and water resource management have been implemented, to control erosion and enhance agricultural productivity and resilience against environmental challenges (Government of Rwanda & IUCN, 2022).

The districts of Muhanga, Ngororero, and Gakenke are the most exposed, experiencing annual soil losses of 46 tonnes, 45 tonnes, and 33 tonnes per hectare, respectively (Government of Rwanda & IUCN, 2022). Muhanga district is further ranked as the second-highest erosion risk area, encompassing 53,352 hectares, which accounts for 82% of the district's total area, where the Nile Nyabarongo Upper Catchment (NNYU) experiences the most severe erosion, affecting 45,961 hectares (Government of Rwanda & IUCN, 2022).

Another previous study in 2016 conducted a sediment fingerprinting analysis of the NNYU to identify potential hotspots of soil erosion. According to the study, the upstream areas of Nyagako sub-catchment were identified as one of the most erosion-prone areas. Moreover, the study recommended further research in the areas that were identified as potential hotspots, where the next step should be to validate the sediment fingerprinting result, confirm the erosion contribution, and identify the causes (GLOWS-FIU, 2016).

1.2. Aim & Objectives

The general aim of this master thesis is to investigate the sources and contributing factors of soil erosion within the Nyagako sub-catchment area and develop an efficient methodology for targeting erosion-prone areas. The purpose of this aim is to facilitate strategic planning for soil conservation investments and other mitigation actions against soil erosion.

The objectives are the following:

- **Identify Erosion-Prone Areas:** Identify areas within the Nyagako sub catchment that are most susceptible to soil erosion, aiming to prioritize them for further investigation.

- **Assess Contributing Factors to Soil Erosion:** Investigate the influence of physical, biological and anthropogenic factors on soil erosion rates within the Nyagako sub catchment.

- **Evaluate Methodological Efficacy:** Analyze the contribution fractions and limitations of the sediment fingerprinting method, in accurately capturing soil erosion processes, further compare with the measuring of turbidity and previous studies, aiming to identify methodological improvement.

1.3. Thesis Outline

This master thesis project was conducted in the Nyagako sub-catchment in Muhanga District and at the University of Rwanda in Kigali, Rwanda in 2023. The report begins with a presentation of soil erosion dynamics and influencing factors, followed by an overview of the methodology, including the links and combinations between the methods and objectives.

The methods used in this study to identify erosion-prone areas within the Nyagako sub-catchment utilizes GIS analysis, sediment fingerprinting analysis, and turbidity measurements. The contributing factors to soil erosion are assessed by examining the topographic and geographic characteristics of the study site through GIS, complemented by fieldwork observations. The potential causes and correlations are analyzed by comparing the results from the different methods and the potential drivers. Lastly, the efficacy of the methodologies used are discussed. The report concludes with a summary of the key findings and recommendations.

1.4. Limitations

The primary limitation of this study lies in the difficulty of accurately measuring and converting the most relevant contributing factors into a model that provides reliable interpretations. Where several constraints have impacted the results, including a limited budget, a restricted timeframe, and the challenging environment in which the project was conducted. These factors have made the execution of the comprehensive and mixed-method approach demanding. Despite this, the project's methodology and analysis are comprehensive, employing both straightforward analysis and advanced mixing models to assess soil erosion.

All the analyses used in this study come with inherent limitations and uncertainties that must be considered when interpreting the results. For instance, the GIS analysis suffers from limited information about the data collection and processing methods executed, however, University of Rwanda is a reliable source where similar study with the same data is used and

was thereby used in this study. The turbidity analysis was hindered by some losses of samples due to pore collection practice, which led to a smaller dataset and increasing uncertainties during the analysis. Similarly, the fingerprinting analysis faced limitations due to the restricted amount of collected data and measurements. Additionally, the sample preparation phase was not performed under optimal conditions, increasing the potential for contamination.

2. Background on Soil Erosion

2.1. Soil Erosion Dynamics

Soil erosion is loosely defined as the transformation of soil into suspended sediment. It is a natural process that occurs without any anthropogenic activities (Svoray, 2022). Soil erosion occurs in a dual-phase process that involves the separation of single soil particles from the larger soil body and their movement by erosive forces like water flow and wind. A third phase, known as deposition, takes place when there is no longer enough energy to carry these particles further (Morgan, 2005).

Water-induced soil erosion can be categorized into several types: sheet erosion, caused by the impact of raindrops and surface runoff forces; gully erosion, characterized by small channels typically 15 to 20 cm deep; channel erosion, affecting the banks and beds of streams; flood erosion, resulting from the flow of floodwaters across plains; and mass movement, encompassing landslides, slope failures, and avalanches (Jain & Singh, 2003).

Rainsplash stands as the primary agent for soil detachment. The impact of raindrops on an exposed soil surface can displace soil particles, propelling them through the air for several centimeters. Repeated exposure to heavy rainfall significantly undermines the soil's integrity (Morgan, 2005). Raindrop splash erosion involves the detachment, transport, and accumulation of soil particles upon impact, primarily driven by raindrops on hillslopes. Studies suggest that slope and wind intensify splash erosion and transported distance (Marzen & Iserloh, 2021). The intensity of storms and rainfall is expected to rise due to global warming, potentially hastening the natural erosion rates in catchment areas (Jain & Singh, 2003). Additionally, the soil undergoes fragmentation through weathering, both from physical changes like wetting and drying cycles, freeze-thaw actions, and frost, as well as through biochemical reactions (Morgan, 2005).

Julien (2010) argues that human actions often speed up soil erosion. For example, agricultural activities, such as plowing and tilling practices makes the soil more prone to erosion. The removal, cutting, or burning of vegetation diminishes the protective ground cover which enhances the erodibility of the soil. Harmful sediment is not the only consequence, but also significantly harms agricultural land by decreasing the fertility of the soil. In certain situations, the rate of erosion can exceed natural geological erosion rates by a factor of 100 to 1,000 times (Julien, 2010).

Furthermore, Jain and Singh (2003) describes that unregulated deforestation, forest fires, overgrazing, inadequate tillage methods, and unsustainable agricultural and land management practices expedite soil erosion, leading to a significant rise in sediment input into waterways. The accumulation of sediment in river channels or reservoirs poses numerous challenges, including elevated streambeds, heightened flood risks, obstruction of navigation routes, and reduced storage capacity in reservoirs (Jain & Singh, 2003). Mining activities can also significantly increase sediment levels in natural waterways. Sediment from mine dumps and spoil banks can erode into nearby streams long after mining has stopped, due to natural rainfall. Additionally, mining gravel from streams can lead to serious channel disturbances, including head cutting upstream, which can cause stability concerns for nearby roads and bridges (Julien, 2010).

Suspended sediment stands as one of the most significant and intrusive pollutants in water systems, affecting water quality through factors like turbidity and serving as a carrier for chemicals and other pollutants. Fine sediment also harms infrastructure, such as dams and turbines, and disrupts aquatic habitats, including areas critical for the spawning (Owen et al., 2016). When a dam is built, it alters the flow dynamics and sediment transport capacity significantly. The wider reservoir slows down the flow velocity and reduces turbulence, leading to sediment deposition. Larger particles and bed load settle near the reservoir entrance, while smaller particles may remain suspended or pass through the dam. This sediment deposition reduces reservoir storage, disrupts water supply, irrigation, and harms

ecosystems of the river by burying vegetation and affecting biotic life due to reduced water clarity and sunlight penetration (Jain & Singh, 2003).

Rill Erosion

The formation of rills is understood to occur where surface runoff consolidates into small flow paths often because of small surface variations, plant cover, or animal tracks on natural slopes, and due to plowing on farmlands (Bryan, 2000; Morgan, 2005). Where flow paths merge, intensifying erosion and forming narrow flow channels, which increases the flow, features such as vortices and turbulence emerge, indicating increasing water velocity from subcritical to supercritical states (Morgan, 2005). This progression is gradual, influenced by factors like soil particle size and sediment concentration in the water, rather than a sudden shift at a specific Froude number threshold (Morgan, 2005).

Gully Erosion

Gully erosion involves the formation of relatively permanent, steep-sided watercourses that experience temporal flows during rainstorms. These formations exhibit erratic behavior with larger sediment loads and rapid changes in slope, often associated with accelerated erosion and landscape instability. Gully formation is initially starting with small depressions due to localized weakening of vegetation cover, which then enlarges into channels as water concentrates and erodes the soil. This process involves both surface erosion and subsurface flow, contributing to significant landscape changes. Gullies can also result from the collapse of subsurface pipes or tunnels, especially in environments with steep hydraulic gradients and soils of high infiltration capacity. The erosive power of gullies is considerable, but their impact varies widely depending on local conditions and storm characteristics, making their contribution to total soil loss difficult to predict without specific local data (Morgan, 2005).

2.2. Factors Influencing Soil Erosion

Rainfall

Soil erosion is significantly influenced by rainfall, primarily due to the detaching power of raindrops hitting the soil surface and the contribution of runoff. The intensity of rainfall is a crucial factor, particularly for erosion by overland flow and rills (Morgan, 2005). High-intensity rainfall leads to increased runoff and the duration of rainfall significantly surpassing the soil's infiltration capacity and enhancing erosion (FAO, 2019). Rainfall erosivity (R) is the estimated capacity of a rainfall to cause soil erosion (McGehee, Flanagan, Srivastava, & Nearing, 2021) and a key factor in the RUSLE model (Zeghmar et al., 2024). The RUSLE model will be further explained in the methodology.

Vegetation

Vegetation reduces the impact of raindrops on the soil surface, enhances soil structure through root growth, and increases water infiltration, all of which reduce erosion risk (FAO, 2019). The plant cover plays a vital role in dissipating the energy of running water by adding roughness to the flow, which reduces its velocity. The level of roughness associated with different types of vegetation depending on factors such as plant morphology, density, and height relative to the flow depth (Morgan, 2005). The RUSLE model assesses this as the cover management factor (C) and several studies show various methods employed to calculate and quantify the factor (Zeghmar et al., 2024).

Topography

Slope length and steepness directly influence the velocity of runoff and its erosive power. Steeper and longer slopes are more susceptible to erosion (FAO, 2019). Moreover, while raindrops scatter soil particles in various directions on flat surfaces, the tendency shifts on sloping terrain, where more soil is dispersed downslope than upslope, a tendency that amplifies with steeper slopes (Morgan, 2005). The LS factor, which combines the effects of

slope length (L) and slope steepness (S), is a critical determinant of soil loss within the RUSLE model (Zeghmar et al., 2024).

Soil characteristics

Soil characteristics, including texture, structure, organic matter, and permeability, critically impact its vulnerability to erosion. Specifically, soils rich in sand, lacking in organic matter, or with compromised structure are highly susceptible to erosion. These properties play a pivotal role in the generation of runoff and the erodibility of soil and the soil's ability to absorb water, with texture being a primary determinant. The susceptibility of soil to erosion is significantly determined by particle size and the stability of soil aggregates; silt-dominated and loamy soils face higher erosion risk, whereas clay and sand-rich soils are more erosion-resistant due to their cohesiveness and particle size. Moreover, the resilience and integrity of soil aggregates, affected by clay content and organic matter, are essential for erosion resistance, as they enhance surface roughness and reduce erosion potential. Both natural and anthropogenic activities modify soil aggregate condition and surface texture, altering erosion susceptibility over time. Although the RUSLE's soil-erodibility factor (K) attempts to consolidate these aspects, it is critiqued for not fully capturing the complexities of rill formation and erosion in diverse landscapes (FAO, 2019).

Soil conservation

In agriculture, practices such as crop rotation, cover cropping, and conservation tillage can significantly reduce soil loss by maintaining soil structure, improving water infiltration, and reducing runoff. These practices help to protect the soil surface from the impact of raindrops and reduce the velocity of surface runoff, thereby minimizing erosion. Engineering measures, on the other hand, include structural interventions like terracing, contour plowing, and the construction of check dams and retention basins. These measures are designed to physically alter the landscape to slow down water flow, encourage infiltration, and capture sediment before it can be transported off-site (Morgan, 2005). The P factor represents the effect of land

use and farming practices on soil erosion. Soils with robust vegetation cover are better protected, mitigating climatic impacts. Additionally, it considers practices such as contour farming and strip cropping, which reduce soil erosion risk by managing water runoff and sediment movement on slopes (Zeghmar et al., 2024).

3. Methodology

This chapter will present the methodology. First, by giving a review of the study site and its geological and topographical features – potential drivers of soil erosion. Next, the development of the sediment fingerprinting analysis will be presented, followed by the fieldwork and turbidity analysis, and finally, how the analysis of potential causes of soil erosion will be addressed. An overview of the methodology is presented in figure 3.1.

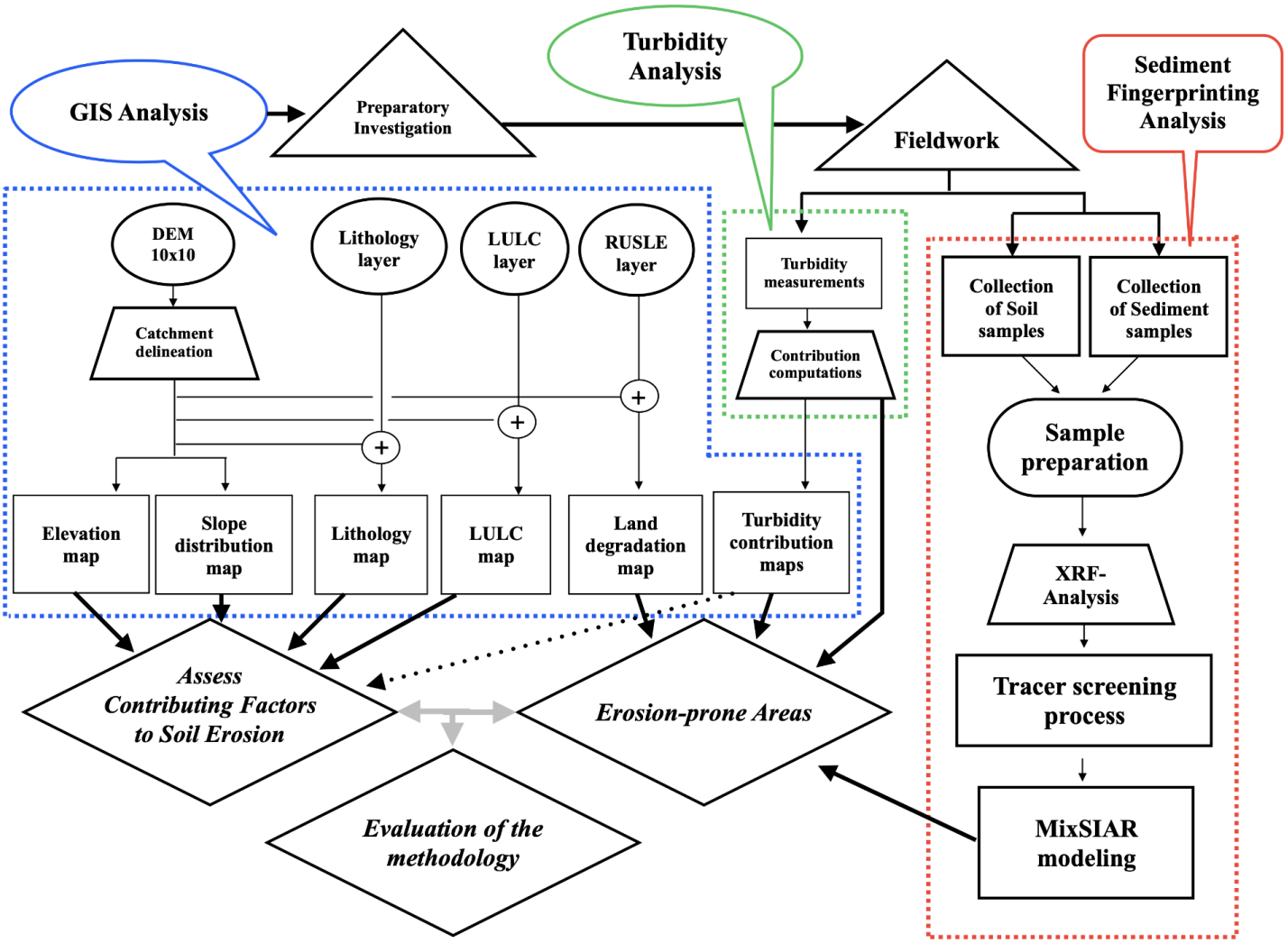

Figure 3.1. Overview of the methodology.

The methodology consists of three major analyses; GIS-, turbidity- and sediment fingerprinting analysis. Both the turbidity and sediment fingerprinting analysis include fieldwork – measurements and collection of samples. Prior to the fieldwork, preparatory investigation was mainly substantiated through GIS analysis. However, the GIS analysis is utilized throughout the project, interpreting and validating the modeling results.

3.1. Study Site Description

The project was conducted in Nyagako sub catchment of the Nile Nyabarongo Upper Catchment (NNYU), which is part of the southern province of Rwanda, see figure 3.2. It is shared by the Muhanga, Mushirio, Nyarusange and Nyamabuye Sectors in the Muhanga District.

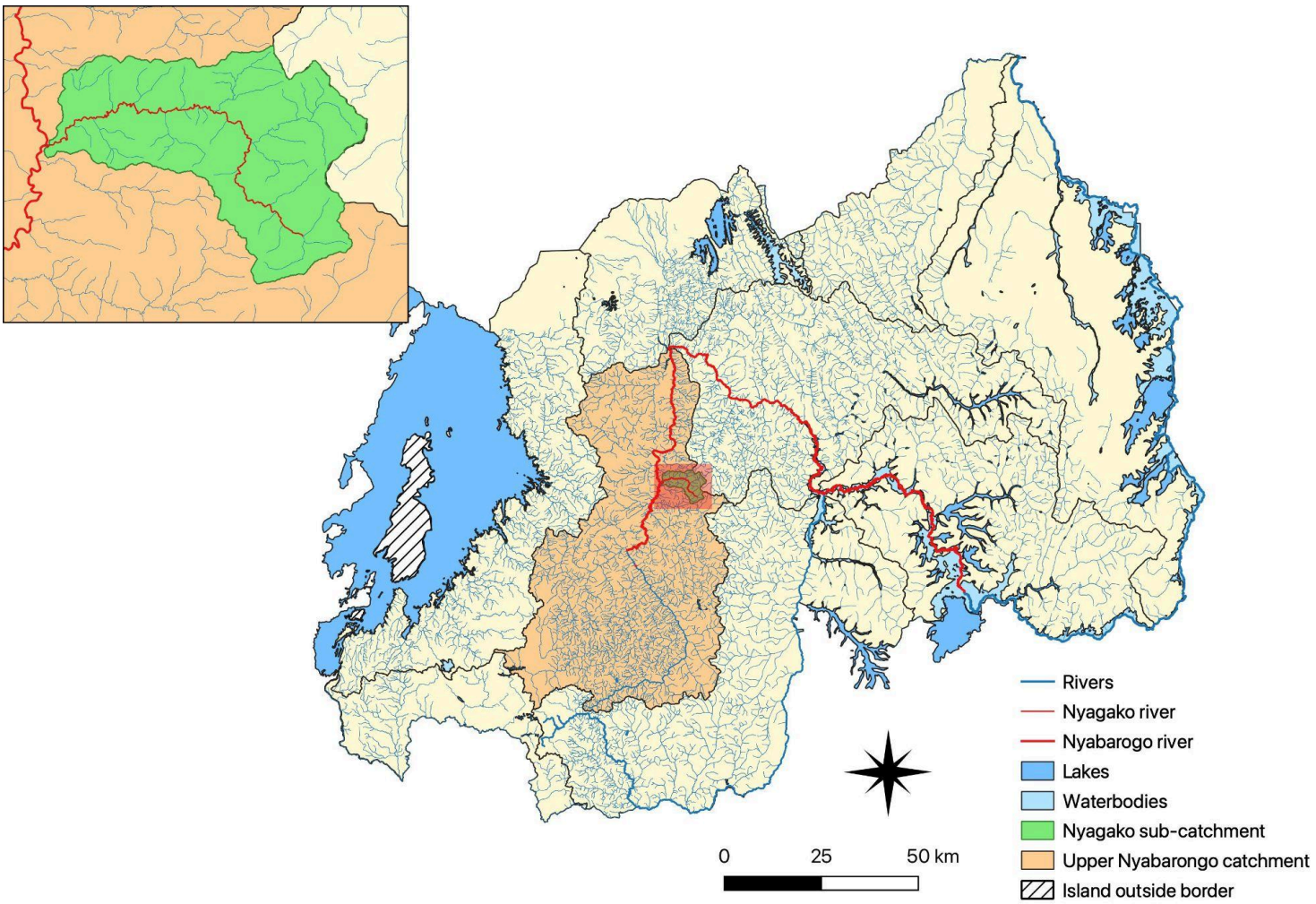

Figure 3.2 Water Bodies in Rwanda. The smaller figure is an enlargement of the Nyagako sub-catchment, indicated by the red highlighted area. Created by author, 2023 (Source: University of Rwanda).

Delimitation, zone division and sampling points

The full extent of Nyagako sub-catchment was not included in the project due to limited access and resources; instead, the most upstream regions were prioritized, which was highlighted as erosion-prone hotspots in the study by GLOWS-FIU (2016). The study site was segmented into three zones: A, B, and C. Zone A (1517 hectares) is the uppermost region of the catchment, Zone B (1148 hectares) is in the mid-part, and Zone C (1589 hectares) is in the downstream end, see figure 3.3.

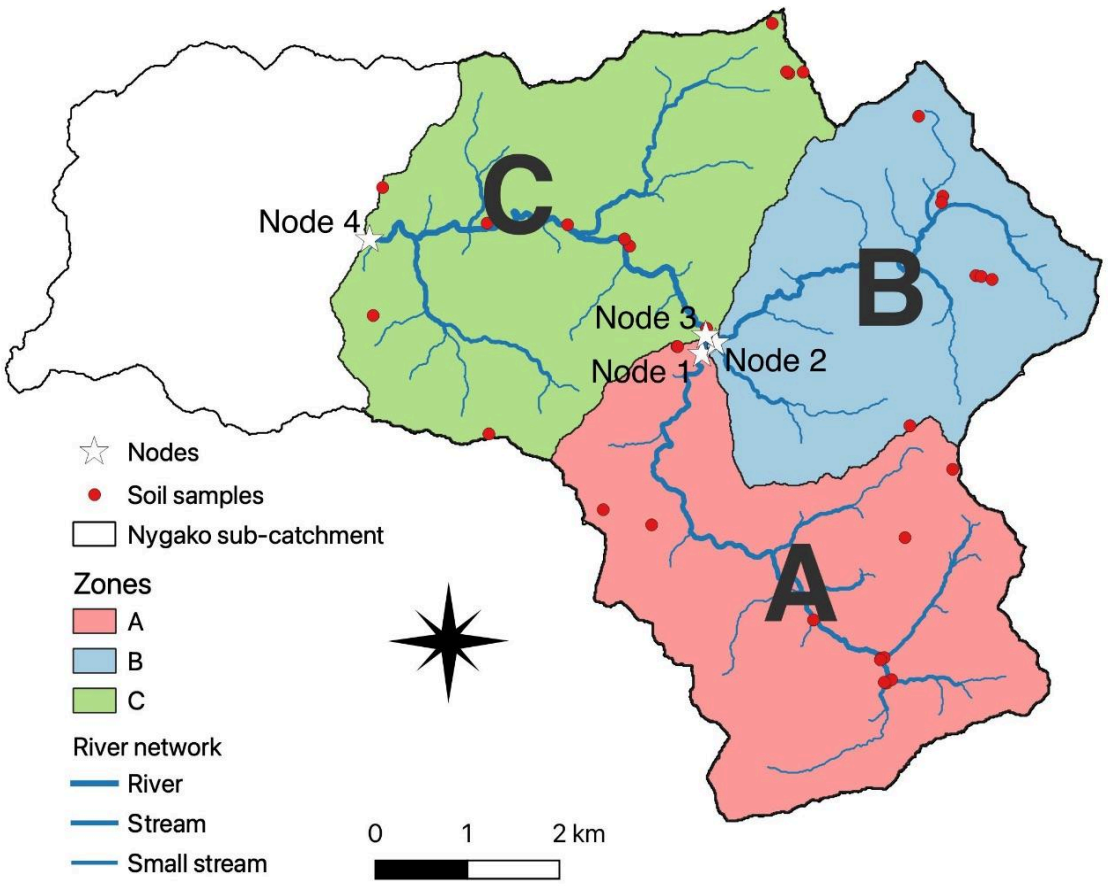

Figure 3.3 Study site, zones and sample locations. Created by author, 2023 (Source: University of Rwanda).

The segmentation identified four key locations – zone outlets and meeting points – referred to in the project as nodes, where sediment samples were collected. Node 1, Node 2, and Node 4 represent the sediment sample locations from Zone A, Zone B, and Zone C, respectively. Node 3 represents the sediment sample location for the mixture of Zone A and Zone B. Location of soil samples are presented in the figure with red dots.

Catchment delineation

The catchment delineation for the study area was conducted as demonstrated in the catchment delineation procedures outlined by van der Kwast and Menke (2020) in QGIS for

Hydrological Applications: Recipes for Catchment Hydrology and Water Management. A high-resolution 10x10m Digital Elevation Model (DEM) was used. The dataset, managed by the Geospatial Data Abstraction Library (GDAL), was provided by the University of Rwanda. The last modification date is recorded as March 12, 2010. The initial step involved reprojecting the DEM to match the Coordinate Reference System (CRS) of the project, ensuring data consistency and spatial accuracy. Once the DEM was reprojected, it was then subset using the 'Clip Raster by Extent' function found under 'Raster extraction', trimming the DEM to the precise area of interest. Following this, the refined DEM underwent a sink filling and spike removal process to correct for any data anomalies. This was accomplished using the 'Fill sinks (Wang & Liu)' tool within the SAGA 'Terrain Analysis - Hydrology' module, which is instrumental in preparing the terrain model for hydrological analysis. The next phase involved the derivation of stream networks through the 'Strahler order' tool, also part of SAGA's 'Terrain Analysis - Channels' suite. This process was done for understanding the flow patterns and stream hierarchy within the catchment. The final step of catchment delineation utilized SAGA's 'Upslope Area' tool, again within the 'Terrain Analysis - Hydrology' module.

Topography

The topography of the Nyagako sub-catchment reveals a distinctly hilly landscape. Elevations within the catchment exhibit considerable variation, ranging from 1474 to 2184 meters above sea level. This variation in elevation is adding a dynamic aspect to the catchment's hydrology, prominently featuring the Nyagako river network, which meanders through the terrain, shapes the catchment's geomorphology and influences sediment transport, see figure 3.4.

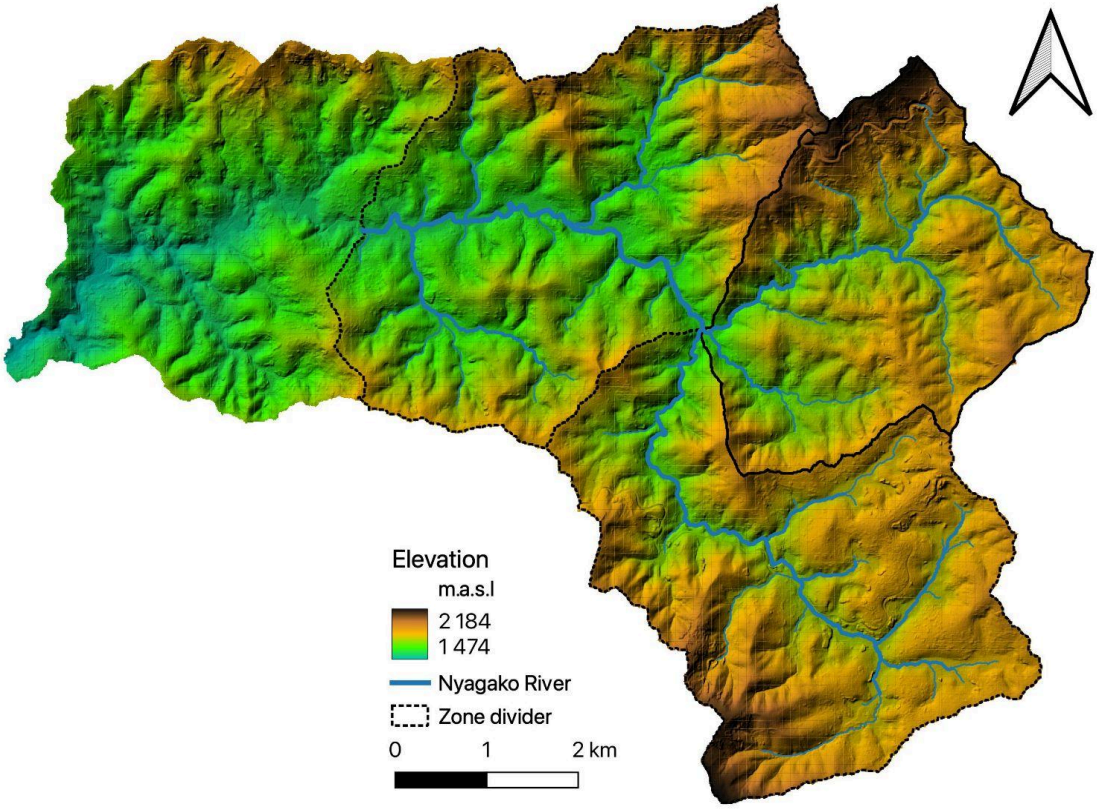

Figure 3.4 Elevation map of Nyagako sub-catchment. Created by author, 2023 (Source: University of Rwanda).

Lithology

The lithology of the study site encompasses three lithological units, denoted as CA, GG, and QS, see figure 3.5. These will be used as the sources for the fingerprinting analysis. CA represents undifferentiated colluvium and alluvial materials, while GG denotes acidic rocks such as granite, gneiss, tonalite, and areas rich in pegmatite. QS comprises quartzite and shale formations, including schists, argillites, ardoises, and phyllites. The site is predominantly characterized by QS (56.5% of the total area), followed by GG (36%), and CA (7.5%).

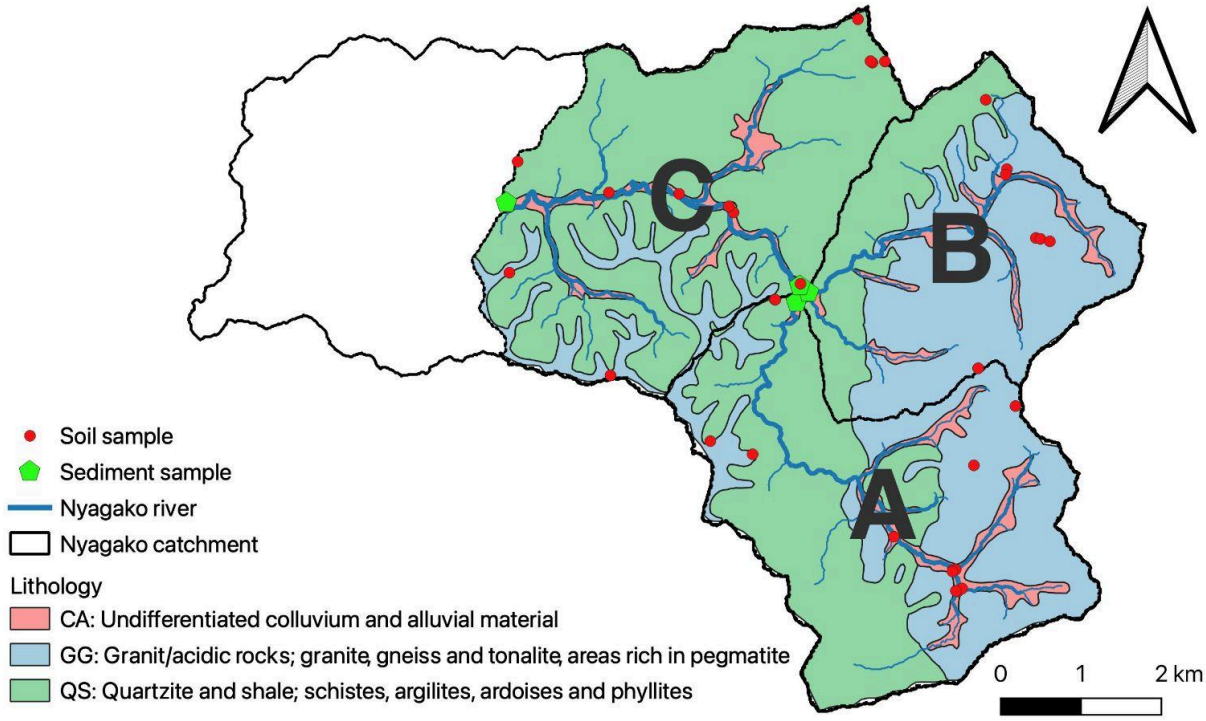

Figure 3.5 Lithology map of the study site, including zones and sample sites. Soil samples represented with red dots and sediment samples with green heptagons. Created by author, 2023 (Source: University of Rwanda).

Land use and land cover

The land use and land cover (LULC) of the study site are predominantly agricultural, accounting for approximately 67% of the land use. Agricultural practices are diverse, including irrigation, terraces, and closed agriculture. Forest cover comprises 24% of the area, with a significant presence of plantation forestry, which may be managed for timber production and erosion control. Additionally, 9% of the land is non-cultivated, encompassing open lands with natural vegetation, built-up areas, and a landfill. The map in figure 3.6, with its diverse color palette, delineates each land use category, offering an overview of how the catchment's land is utilized and managed.

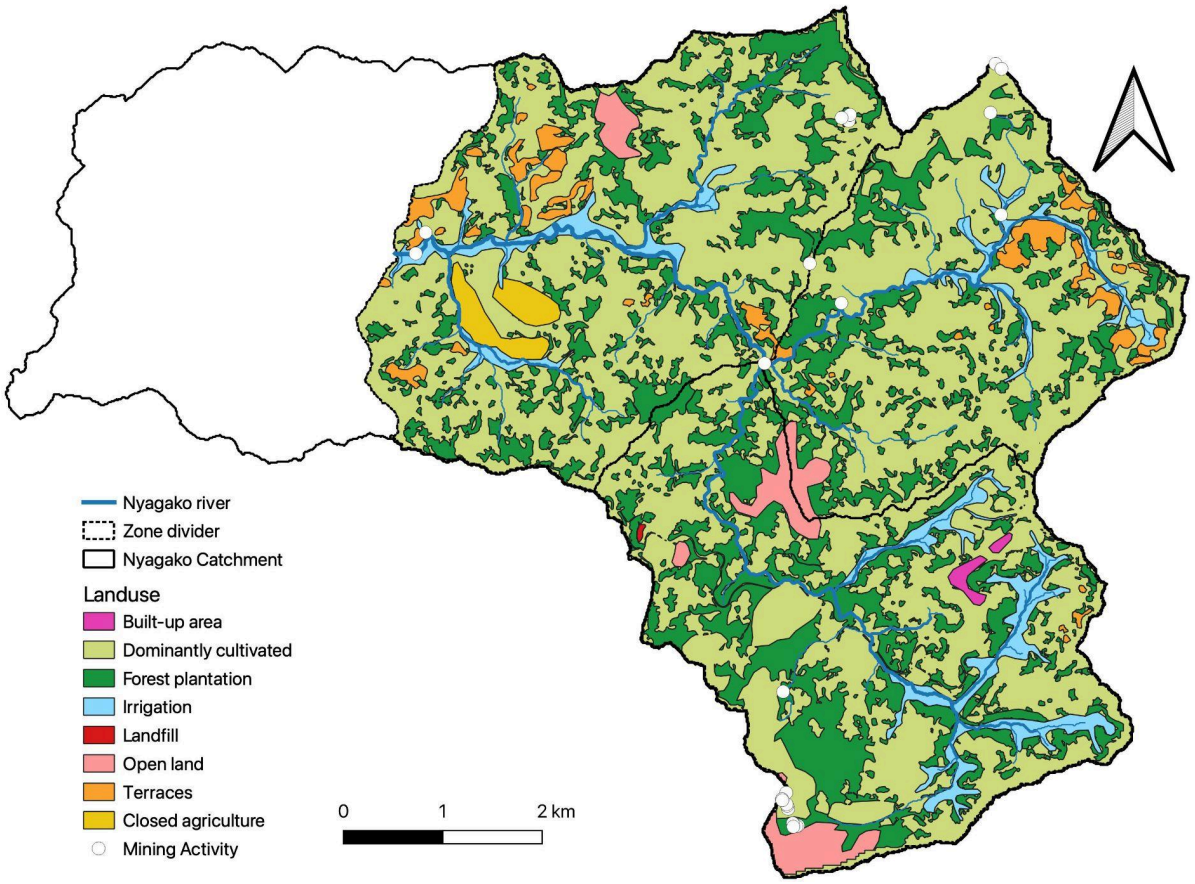

Figure 3.6 Land use and land cover map of the study site in Nyagako sub-catchment. Base map provided by University of Rwanda and location of landfill, terraces, and mining activity added by the author.

Mining activities are identified at 23 locations across the catchment. These sites vary in their operational status and legality, ranging from officially sanctioned mining operations to unauthorized or informal extractions. Such activities can significantly impact the turbidity and total suspended load within the river, altering hydrological systems and potentially affecting water quality.

The LULC distributions were analyzed for each source (lithology type) in the three zones; A-C. LULC was segmented into three main categories: Forest, Cultivated, and Non-cultivated, as defined by the LULC map. Specifically, 'Cultivated' encompassed areas

designated for irrigation, terraces, and closed agriculture; 'Forest' was defined by regions of forest plantation; and 'Non-cultivated' included open land, built-up areas, and landfills. This tripartite classification facilitated a detailed investigation of land use dynamics on soil erosion to analyze possible causes of soil erosion.

Slope distribution: mapping and categorization

The slope distribution map was created by using the DEM (10x10m) raster layer. The 'Slope' function from the 'Raster terrain analysis' toolset calculated the terrain gradient. Then, the 'Reclassify by table' option under 'Raster analysis' reclassified the resulting slope values. The categorization was set according to the slope angle classifications from the study 'USLE-Based Assessment of Soil Erosion by Water in the Nyabarongo River Catchment, Rwanda' by Karamage et al., (2016), which identified slopes as 'Very Gentle to Flat' (<5°), 'Gentle' (5°-<15°), 'Steep' (15°-<30°), and 'Very Steep' (>30°).

Revised Universal Soil Loss Equation (RUSLE)

RUSLE is an established method to assess and identify soil erosion-prone zones within a GIS environment. It utilizes five key factors: rainfall erosivity, soil erodibility, topography, cover management, and conservation practice, the model is empirical and further used to estimate the annual average soil loss (Zeghmar et al., 2024). For this study, data based on RUSLE from the Rwanda Land Degradation Monitoring and Assessment 2022 dataset, created by PKabatha (2023) was used as a complement to validate the MixSIAR and turbidity result.

3.2. Sediment Fingerprinting Analysis

The fundamental concept of sediment fingerprinting is that the characteristics of the sediment, such as concentrations of geochemical elements or activity of radionuclides, indicate its source (Owens et al,. 2016). It has been widely adopted across various fields to study numerous catchments around the world since its early development in the 1970s and 1980s

(Smith, Karam, & Lennard, 2018), employing a diverse array of physical and biogeochemical tracers at different catchment scales (GLOWS-FIU, 2016). The global challenges due to soil erosion has increased the development of geochemical, radiochemical and isotopic fingerprinting techniques (Blake et al,. 2018). Utilizing sediment fingerprinting methods, alongside data on sediment transport and budgeting, can yield significant insights into landscape dynamics and offer valuable information that can inform the management of river basins and coastal areas (Owens et al., 2016).

The use of geomorphological information is valuable for selecting the tracing approach. In catchments with heterogeneous lithologies, i.e. diverse rock compositions, a strategy based on geochemical element concentrations would be reasonable for distinguishing between sources, corresponding to specific geological signatures, including features as steepness in the upper regions with hard rock versus hillsides composed of softer, more easily weathered materials (Evrard et al., 2022). It is recommended to keep the number of sources for discrimination to a minimum, with a guideline suggesting a cap at four sources (Evrard et al., 2022; Lees, 1997). Typically, un-mixing models require at least (n – 1) distinct tracers for a contribution from n sources in a mixture. Ideally, every source should be uniquely identified by at least one tracer that effectively distinguishes it from the others (Evrard et al., 2022). Thereby, the choice of sources for this study is based on the three lithology units in the catchment, where geochemical elements are used as tracers.

Tracer screening process

The tracer selection is an important and critical step for a successful estimation of the source contribution in the sediment (Chalaux-Clergue et al., 2024) and the accuracy can be significantly influenced by the selection of tracers (Smith, Karam, & Lennard, 2018). Sediment fingerprinting relies on tracers that remain unchanged during transport and that can be statistically distinguished between sources, ensuring accurate identification of sediment origins while excluding environmentally variable elements (Wilkinson et al., 2013).

Tracers that remain unchanged during transport from source areas to deposition sites are considered as conservative (Smith, Karam, & Lennard, 2018). Non-conservative behavior of a tracer is primarily assigned with two key phenomena. Firstly, there's the potential for particle size sorting along the transportation route, influenced by various factors such as the extent of runoff, the severity of rainfall, the flow of the river (Chalaux-Clergue et al., 2024). To mitigate the effects of particle size sorting on the characteristics of sediment, the fraction < 63 μm is typically isolated by sieving of both the sources and the sediment sample for analysis (Evrard et al., 2022). According to the Wentworth scale, the sediment includes colloid (0 - 1 μm), clay (0 - 3.9 μm) and silt (3.9 - 62.5 μm) (Haldar & Tišljar, 2014). Secondly, biogeochemical activities can occur as particles are transported. It is influenced by the extent the tracers undergo biogeochemical processes, including dissolution, adsorption, oxidation, and reduction (Chalaux-Clergue et al., 2024). For this study, only samples with fractions < 63 μm were used.

To evaluate the conservative nature of a tracer, the most conventional approach is by comparing the range of source soil samples concentration to the sediment samples mean concentration (Wilkinson et al., 2013). This is often referred to as a “range test”, which is often analyzed through generated box plots for upstream sources and their corresponding sediment sample across all tracers, with the mean sediment concentration evaluated to determine whether they are predominantly aligned within the range of the mean concentration of the upstream sources (James et al., 2023). Any sediment’s tracers’ mean concentration that did not align within the source range shall be excluded. Although, it's important to note that range tests, by their nature, do not provide a quantitative confirmation of the complete absence of non-conservative behavior (Chalaux-Clergue et al., 2024).

A range test was performed in this study to assess the conservative behavior of each potential tracer element after the raw geochemical dataset was examined and sorted elements in order that showed high concentration performance and the samples were also sorted into potential source groups, based on the geographical information and field observations, such as sample

location and lithology, in appendix A1. The elements in table A1.1 were excluded from further analysis because they had a high degree of ‘non-detected values’, denoted ‘ND’, suggesting that the actual concentration is unknown of that element. According to other studies (James et al. 2023), tracer elements that deviated too much from the expected range or clearly had poor performance are excluded from further analysis.

A common method to assess the tracer's ability to differentiate among the sources (discrimination capacity), is the non-parametric Kruskal–Wallis H test. The outcome of this test reveals whether there is a significant difference between at least one group and others and helps in identifying tracers that can effectively differentiate between various sources (Chalaux-Clergue et al., 2024; Evrard et al., 2022). Any significant differences between the elements, where all tracer elements that had a p-value larger than 0.05, should be removed from further analysis. Another method is to check the tracer if it has higher intrasource variance than intersource variance, if so, it should be removed (James et al. 2023). Both of these methods were applied in this study.

A Shapiro-Wilk test was applied in the screening process to remove tracers that were not normally distributed, in alignment with the model's presumption that mixture tracer data follow a normal distribution, as stated by Stock et al. (2018).

After assessing conservative behavior and discriminant power, it is common to conduct a principal component analysis (PCA), to provide an overview of the variance in the tracers and samples (Chalaux-Clergue et al., 2024). PCA can be used to simplify complex datasets by examining variances within and between sources, enabling the identification of distinct geochemical sources (James et al., 2023). A PCA was performed in this study.

Model development

A Bayesian mixing model was developed with the R package MixSIAR (Stock et al. 2018) to evaluate the contribution of sources in mixtures through tracer elements. Based on observed

tracers in the soil and sediment samples, the MixSIAR model generates expected contributions from different sources associated to each node, through so-called posterior proportion contributions.

Bayesian mixing models were first developed as a statistical tool in ecological studies to estimate diets of animals (Hopkins and Ferguson, 2012) by using stable isotopes as tracers to evaluate the proportion of various prey (sources) in the diet of a consumer (mixture). However, these models are also applied in other research areas, including tracking animal migration, identifying sources of pollutants, tracing nutrient transfers between ecosystems and sediment erosion by fingerprinting analysis (Stock et al., 2018). These models are based on Bayesian inference, a statistical approach that updates the probability of hypotheses based on new evidence. This process incorporates prior knowledge or beliefs with new data, enabling the estimation of probability distributions on model parameters, treating unknown parameters and observable data as random variables, and computing the distribution of these parameters conditioned on the observed data (Prieto Tejedor, 2017). In the context of mixing models, Bayesian methods estimate how much each potential source contributes to a mixture by considering prior knowledge about the sources and the observed data from the mixture (Blake et al., 2018).

R software is a powerful tool for conducting Bayesian data analysis, especially for complex models. It is highly valued in the Bayesian community for its versatility, extensive resources, and being free of charge. Bayesian Markov Chain Monte Carlo packages like JAGS are integrated within R, enhancing its utility in Bayesian data analysis (Kruschke, 2015). For this study, the tracer screening process and MixSIAR modeling were performed using R version 4.3.1, R Core Team, (2023).

Markov Chain Monte Carlo (MCMC) is an algorithm used for estimating complex statistical models through simulation of random samples. It operates akin to a guided exploration, making decisions at each step by evaluating the likelihood of moving to a new

position versus staying at the current one. **JAGS**, which stands for Just Another Gibbs Sampler, is a specific program designed to implement these MCMC methods. It allows users to efficiently conduct Bayesian data analysis by building and running MCMC samplers for complex hierarchical models. The integration of JAGS within R, facilitated by packages like rjags and runjags, enhances the capacity for Bayesian data analysis, demonstrating the powerful synergy between MCMC methods and computational tools (Kruschke, 2015).

MixSIAR is an open-source R package developed by Stock et al., (2018) to perform Bayesian statistical computations through JAGS for its capabilities of generating MCMC samples, which is essential for estimating the contributions of different sources in a mixing system. One of the main features of MixSIAR according to Stock et al., (2018) is its ability to include both fixed and random effects as covariates, which helps in accounting for variability in mixture proportions (Stock et al., 2018). Additionally, evaluate the relative support for various models by using information criteria (Stock et al., 2018). Through MixSIAR, the developers aim to unify mixing model tools and provide a foundation for future improvements in the analysis of mixing systems (Stock et al., 2018). Although Bayesian mixing models through the MixSIAR framework have been effectively applied in various river basin studies to discern sediment sources, their comparative analysis between mixtures and sources tends to be restricted to specific catchments, with diagnostic precision decreasing as the scale expands (Blake et al., 2018).

To assess the source apportionment in river basins, a hierarchical mixing model approach named **'Deconvolutional MixSIAR' (D-MixSIAR)** has been developed, which expands MixSIAR's capabilities to explicitly consider the 'structural hierarchy' inherent in a river basin or catchment. D-MixSIAR operates by sequentially employing the MixSIAR mixing model to consecutive sediment mixture points (nodes), such as those found downstream of a complex river network. Importantly, within D-MixSIAR, the data contributed as an upstream mixture is organized as a source at a downstream node (Blake et al,. 2018).

Each node was first modeled separately, employing three tracers that were selected based on the tracer screening process. Three sources, labeled CA, GG and QS, were defined based on the lithology in the study site and each has a geochemical composition of the tracer element that passed the tracer screening process. In all MixSIAR model implementations, a residual error term was used, acknowledging the uncertainties in the measurements, and all source contributions were treated as equally likely. The MCMC parameters were typically configured as follows: a chain length of 1000000, a burn-in period of 700000, a thinning interval of 300, and three parallel chains.

Based on the MixSIAR model result, each upstream node was then deconvoluted according to the D-MixSIAR framework, by Blake et al. (2018). This process involved decomposing the posterior proportion contributions from an upstream node into designated source groups. This process enabled an identification of contribution from upstream source to the composition at the downstream nodes. This cumulative aspect underscores the interconnectedness of the catchment's sediment dynamics and emphasizes the importance of considering the catchment as an integrated system when interpreting sediment sources and their contributions. A conceptual overview of the model structure is presented in figure 3.7.

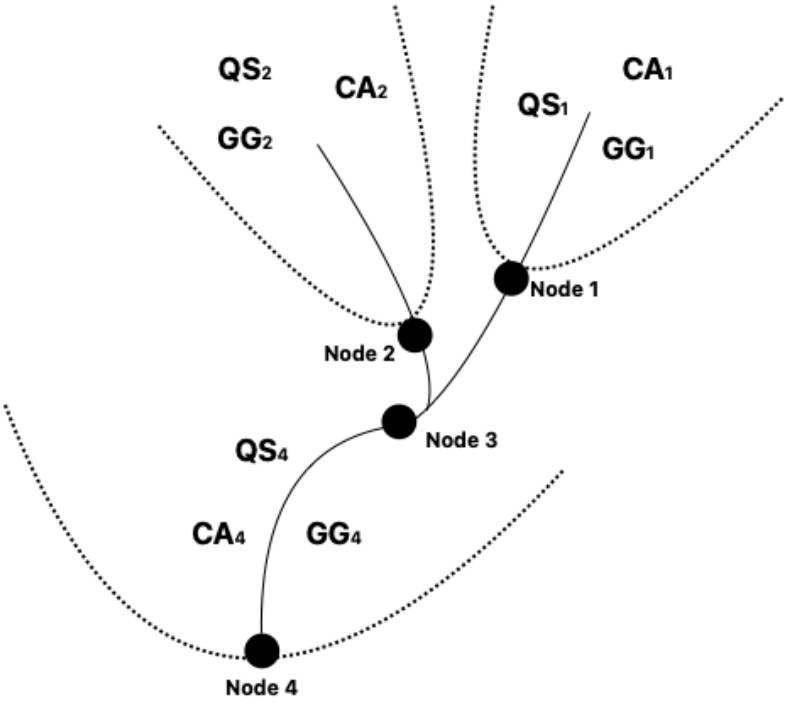

Figure 3.7 Conceptual design of the D-MixSIAR model structure, illustrating the hierarchical mixing process. Node 1-4 represents the progressive stages of mixture. 'CAx', 'GGx' and 'QSx' represent the source groups selected and the corresponding upstream sources. The dashed lines represent sub-catchment boundaries.

At Node 1, which represents the outlet of Zone A, a mixture of upstream sources CA_1, GG_1, and QS_1 is observed. Similarly, Node 2 corresponds to the combined upstream source of CA_2, GG_2, and QS_2. However, Node 3 lacks its own distinct upstream sources of CA, GG, or QS; instead, it relies on Node 1 and Node 2 as additional upstream sources. These sources can be deconvoluted into their respective components (CA, GG, or QS) using the posterior contribution output from MixSIAR. For Node 3, the contributions from Node 1 and Node 2 were calculated as follows in Eq.1:

$$\begin{aligned} CA_3 &= P_{Node1} \cdot CA_1 + P_{Node2} \cdot CA_2 \\ GG_3 &= P_{Node1} \cdot GG_1 + P_{Node2} \cdot GG_2 \\ QS_3 &= P_{Node1} \cdot QS_1 + P_{Node2} \cdot QS_2 \end{aligned} \qquad \text{(Eq.1)}$$

Where, P_{Node1}, P_{Node2}, represent the proportion contribution from Node 1, Node 2 and CA_x,GG_x and QS_x represent the deconvoluted proportion contribution of each source at its node.

Similarly, for Node 4, the contribution from upstream sources and Node 3 were determined as in Eq.2:

$$CA_4' = P_{Node3} \cdot CA_3 + CA_4$$
$$GG_4' = P_{Node3} \cdot GG_3 + GG_4 \quad \text{(Eq.2)}$$
$$QS_4' = P_{Node3} \cdot QS_3 + QS_4$$

Here, P_{Node3} represents the proportion contribution from Node 3 and CA_3, GG_3 and QS_3 are the deconvolved proportion contribution result.

This means that the sediment at Node 4 encapsulates the contribution from all upstream processes, intermixing local inputs (CA_4, GG_4 and QS_4) with those transported from Node 3. The final deconvolved result at Node 4 is CA_4', GG_4' and QS_4'.

The average value and standard deviation of samples representing upstream sources for each respective node were utilized in the MixSIAR modeling, consistent with the sample design and MixSIAR framework. The input data for both the source, mixture and discriminant input in each MixSIAR model is detailed in the appendix A2.

When analyzing the results from the MixSIAR modeling, it is crucial to examine the entire posterior distribution graph. The interpretation of what source is contributing the most relative and depending on what values that are highlighted. The modes, corresponding to the graph's peaks, represent the most frequently occurring contribution fractions and suggest

potential ultimate distribution contributions from the sources. Unlike the mean, the mode is not affected by extreme values, making it a reliable measure of central tendency, especially when the data include outliers that could skew the mean. The median, denoting the central contribution, is particularly useful in distributions that are asymmetrical or include outliers, as it provides a more accurate reflection of the typical value. Meanwhile, the mean calculates an average of all potential contributions modeled. It is essential to recognize the variation and spread of these values, as indicated by the range of the confidence intervals. Generally, the posterior distribution plots show a relatively high standard deviation, which indicates a degree of uncertainty in the results.

A comparative analysis was performed between the proportional contributions derived from the posterior outputs of the MixSIAR models and the estimated turbidity levels at corresponding nodes and the potential land use degradation map (RUSLE). This comparison seeks to align and validate the result.

3.3. Fieldwork

The fieldwork was conducted over six days, spanning from 20th of September to 10th of December 2023. It included two field visits and four sampling campaigns. The field visits aimed to lay the necessary logistical groundwork for ongoing research, assess the accessibility of key locations, and observe the land use and terrain. Observations made during these visits led to adjustments in the project's objectives and identification of sites for the sampling campaigns.

Four campaigns were conducted, three sediment sample campaigns and one soil sample campaign. The aim of the campaigns was to collect data for the sediment fingerprinting analysis according to the D-MixSIAR modeling design. Sediment samples were collected along the river at four nodes during the same day at three different occasions. Measurements of the turbidity from connected streams and the main river were also carried out during the

sediment campaigns. The strategy for the soil sample campaign was to collect soil samples around the study site to get a broad geochemical representation of the catchment, where the primary focus was to collect subsoil, although topsoil was often collected as a complement.

Sampling strategies

Finding a balance in the quantity of samples to gather is essential, considering the limitations of time, budget, and logistical challenges. Nonetheless, it's advantageous to collect as many samples as possible. A greater volume of source samples invariably leads to a stronger foundation for analysis, modeling, and debate (Evrard et al., 2022; Clarke & Minella, 2016). In Addition, the understanding of erosion and sedimentation processes by residents can assist in pinpointing critical areas of erosion and sedimentation, potentially streamlining fieldwork to be more effective with fewer samples required (Evrard et al., 2022). Moreover, the point at which sediment samples are collected during a flood event can significantly influence the outcomes of sediment fingerprinting analyses, given that the origin of sediments passing through catchment exits can change dramatically throughout runoff episodes (Evrard et al., 2022).

Soil sampling

A total of 33 soil samples were collected to ensure a representative sampling of the catchment *(see figure 3.3 in section 3.1 – Study Site Description)*. This collection resulted in 10 samples from the topsoil layer and 23 from the subsoil layer. Notably, the topsoil samples were primarily obtained in conjunction with the subsoil samples or when the subsoil layer was not accessible. This approach was adopted to accommodate the initial uncertainty regarding the optimal sampling strategy and to maintain consistency with the sediment samples. The target was to collect samples that contained representative mineral composition of the location. The soil samples were collected in zip bags that were purchased from a pharmacy in Kigali. The amount of soil collected was determined arbitrary, but in general as

much as the zip bag could contain. Excavated land, gullies and channel banks made it possible to collect subsoil samples, see figure 3.8.

Figure 3.8 Top left: Gully where a soil sample was collected. Top right: Collected soil sample in zip bag from excavated land. Bottom left: Channel bank where a soil sample was collected. Bottom right: Both top and subsoil illustrated.

Sediment sampling

A total of 13 sediment samples were collected, with three samples at each node, except for node 4, where four samples were collected. The sediment samples were collected by using clean PET bottles, where each sample ranged from approximately 4.5 to 5 liters per sample, see figure 3.9.

Figure 3.9 Sediment sampling at Node 3.

Sampling occurred in the middle of the river at a depth of a few centimeters. At least three sediment samples per node were collected, following recommendations from previous studies on sampling of mixtures in relation to the upstream sources (Blake et al,. 2018).

Sample preparation

All samples went through preparation before further analysis.

The soil samples were air dried in an oven at 105° C for at least 24 hours and then sieved through (< 63 µm) pore size, see figure 3.10. Some of the samples had to be gently disaggregated by using a pestle and mortar.

Figure 3.10 Left: Air-dried soil samples. Right: Sieved soil samples through < 63 µm pore size.

The sediment samples were sieved through < 63 µm pore size, then heated on a stove to remove the water. When almost all water had been evaporated, the remainder of the sample was transferred to a plate to be finalized in the oven at 105° C, see figure 3.11. The final sediment sample was then gently removed from the plate by hand using a plastic glove.

Figure 3.11 Left: Evaporation of sediment sample. Right: Evaporated sediment sample.

After the sample preparation, the next step was to analyze all of the samples through a 'Handheld X-ray Fluorescence (XRF) analyzers: X-MET8000 range' at a duration of 60 seconds to get the geochemical composition, see figure 3.12.

Figure 3.12 X-ray Fluorescence (XRF) - analysis.

Turbidity measurements

Turbidity measurements were conducted upstream of the outlet of a sub-catchment, within the stream originating from the sub-catchment, and at points further downstream where mixing had taken place, as indicated by the sub-catchments in figure 3.13.

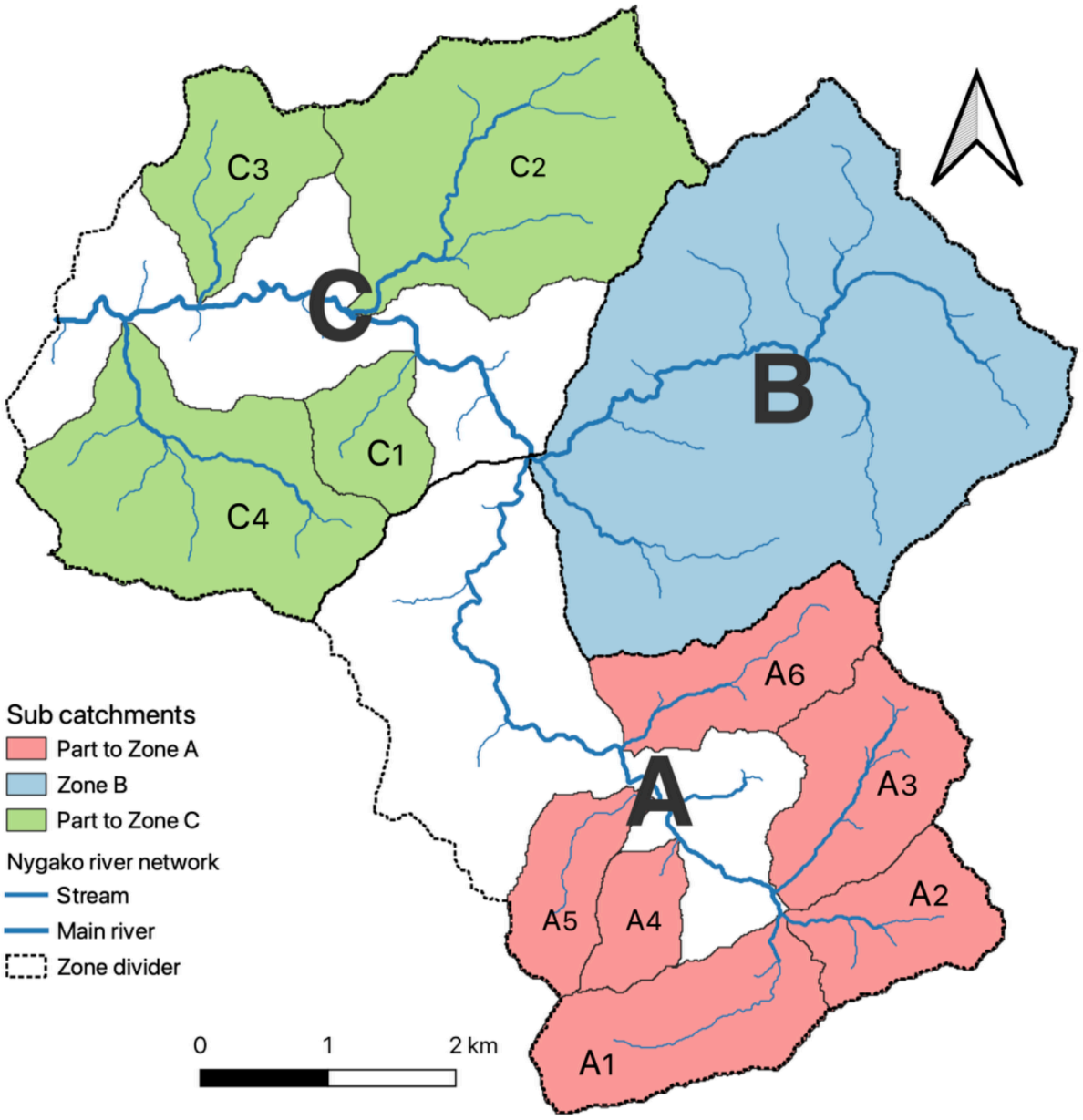

Figure 3.13 Sub-catchments for the turbidity sampling in the study site within Nyagako sub-catchment. Created by author, 2023 (Source: University of Rwanda).

The measurements were conducted using a 2100P ISO Portable turbidimeter, supplied by the University of Rwanda. Prior to measurement, the turbidimeter underwent calibration using distilled water. In cases where sample turbidity exceeded 1000 NTU, dilution was necessary. This involved removing half of the sample and compensating with distilled water. The measurement was either done directly in the field or in the laboratory, see figure 3.14.

Figure 3.14 Measuring turbidity in situ (Nyagako sub-catchment).

Additionally, TSS (total suspended solids) measurements were taken as a complement to the turbidity measurements. The measurements of the TSS showed a linear correlation to the turbidity measurements and are presented in the appendix A1.

To compute turbidity contribution fractions, principles of mass conservation and the mass balance equation was applied, an equation (Eq. 3) was derived, based on the measured turbidity levels:

Where, c is the resulting turbidity level from the mixing of two streams with different turbidity levels, denoted as c_1 and c_2, and flows, Q_1 and Q_2.

$$c = \frac{Q_1 \cdot c_1 + Q_2 \cdot c_2}{Q_1 + Q_2} = \frac{Q_1}{Q_1 + Q_2} \cdot c_1 + \frac{Q_2}{Q_1 + Q_2} \cdot c_2 = \left[\alpha = \frac{Q_1}{Q_1 + Q_2}\right] = \alpha \cdot c_1 + (1 - \alpha) \cdot c_2 \qquad \text{(Eq.3)}$$

This represents a linear combination of the two turbidity levels, weighted by the proportion in which they are mixed. The parameter 'α' is the fraction of the total flow that comes from Q_1, which in (Eq.4) can be expressed as:

$$c - c_2 = \alpha(c_1 - c_2) \Rightarrow \alpha = \frac{c - c_2}{c_1 - c_2} \qquad \text{(Eq.4)}$$

3.4. Potential Causes of Soil Erosion

The analysis of potential causes and correlations to soil erosion was conducted using outputs from GIS analysis, turbidity measurements and computations, and fingerprinting modeling results. The analysis explored the following correlations:

Contribution Per Hectare vs. Average Slope Gradient: Estimations of soil erosion contribution per hectare were plotted against the average slope gradient for each source (CA, GG, and QS) to examine the influence of topography on soil erosion in the study site.

Contribution vs. Cultivation-Forest Ratio Correlation: The relationship between the per-hectare contribution of soil erosion and the ratio of cultivated land to forested land was analyzed to reveal the potential differential contributions to soil erosion based on land use dynamics.

4. Results

The results are presented in the following sequence: First, a review of the results of the contribution models are presented, and analysis of the potential land degradation distribution based on RUSLE. Next, the potential drivers of soil erosion, where LULC distribution is combined with distribution of the lithology and zones, Including, potential impact from mining activity and rainfall events are addressed. Finally, results from the analysis of the potential causes and correlation of soil erosion are presented.

4.1. Contribution Modeling Result

Fingerprinting analysis result

The result from the tracer screening process is summarized in table 4.1 and the result from the PCA is plotted in figure 4.1. A more detailed presentation of the results from the tracer screening process can be found in the appendix A1.

Table 4.1 Summarized results from the tracer screening process from left (start) to right. Elements that passed a test were further analyzed in the next stage. The once that failed was not further analyzed

	Elements of significance		*Conservative behavior*		*Discriminant power*			*Normal distribution*		*Evaluation*
Start →	**XRF result**	→	**Range test**	→	**Kruskal-Wallis test**	**Intrasource vs Intersource**	→	**Shapiro-Wilk test**	→	**PCA**
Passed:	Si, Al, Fe, K, Ti, Ca, P, Zr, Rb, Nb, Mn, Zn and Pb.		Si, Al, Fe, K, P, Rb, Mn, Zn and Pb.		Si, Al, Fe, K and Mn	Si, Al and Fe		Si, Al and Fe		Si, Al and Fe
Failed:	*Ba, Sr, Mg, As, Cu, Sb, Cr, Ta, Co and Ni.*		*Ti, Ca, Zr and Nb*		*P, Rb, Zn and Pb.*	*K and Mn*				

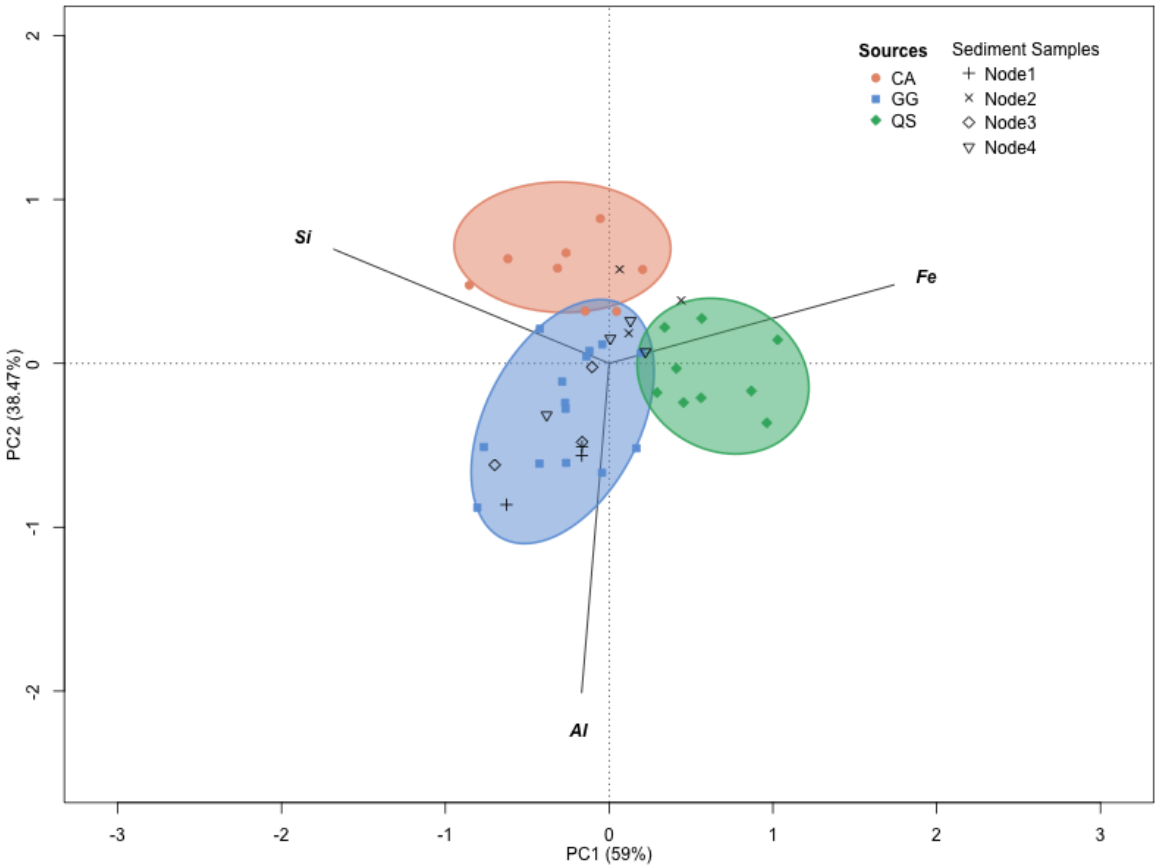

Figure 4.1 Plot of the principal components for tracers Si, Al and Fe, plotted with the sediment samples at nodes(1-4) for comparison. CA: undifferentiated colluvium and alluvial material, GG: Granit/acidic rocks; granite, gneiss and tonalite, areas rich in pegmatite, QS: Quartzite and shale; schistes, argilites, ardoises and phyllites. Ellipses represent 75% of group variability.

The PCA graph illustrates the distribution among the samples, how they are distinguished from each other based on their geochemical signature of the resulting tracers; Si (Silicon), Al (Aluminum) and Fe (Iron). The principal components (PC1 and PC2) explain a significant proportion of the variance in the dataset (59% and 34.87%, respectively). The plot demonstrates groupings among sources (CA, GG and QS) and sediment samples from different nodes (1-4), which may reflect differences in geochemical composition. The arrows representing Si, Al and Fe indicate their influence on the PCA; their direction and length suggest how each contributes to the variance captured by PC1 and PC2. Overlapping ellipses between different sources suggest similar geochemical signatures, while distinct ellipses imply different compositions. The closeness of samples within each ellipse indicates

similarity within that group. CA, GG and QS are three groups that each have distinctly separated geochemical signatures. CA has higher concentrations of Si, while GG and QS have higher Al and Fe respectively.

The output from the MixSIAR modeling are the posterior distribution graphs for each node (1-4), illustrating the relative contribution from the upstream sources (figure 4.2). The densities of these distributions are scaled so that their area sums to one, providing a normalized view of the source contributions. Table 4.2 shows the statistical summary associated with figure 4.2.

Table 4.2 Statistical summary of the MixSIAR result for Node 1-4.

	Source	**Mode**	**Median**	**Mean**	**SD**	**2.5%**	**25%**	**75%**	**97.5%**
Node 1	CA	0.033	0.097	0.135	0.126	0.004	0.430	0.188	0.466
	GG	0.832	0.651	0.609	0.242	0.070	0.449	0.810	0.952
	QS	0.069	0.212	0.256	0.202	0.008	0.092	0.374	0.758
Node 2	CA	0.456	0.397	0.403	0.215	0.031	0.233	0.557	0.832
	GG	0.053	0.160	0.205	0.169	0.005	0.070	0.302	0.622
	QS	0.444	0.390	0.392	0.202	0.037	0.239	0.536	0.789
Node 3	P_{Node1}	0.797	0.733	0.712	0.180	0.269	0.605	0.846	0.982
	P_{Node2}	0.203	0.267	0.288	0.180	0.018	0.154	0.395	0.731
Node 4	CA	0.221	0.228	0.237	0.131	0.020	0.138	0.323	0.659
	GG	0.080	0.203	0.237	0.174	0.011	0.096	0.344	0.659
	P_{Node3}	0.182	0.236	0.260	0.172	0.012	0.122	0.367	0.649
	QS	0.266	0.262	0.265	0.130	0.032	0.0175	0.352	0.535

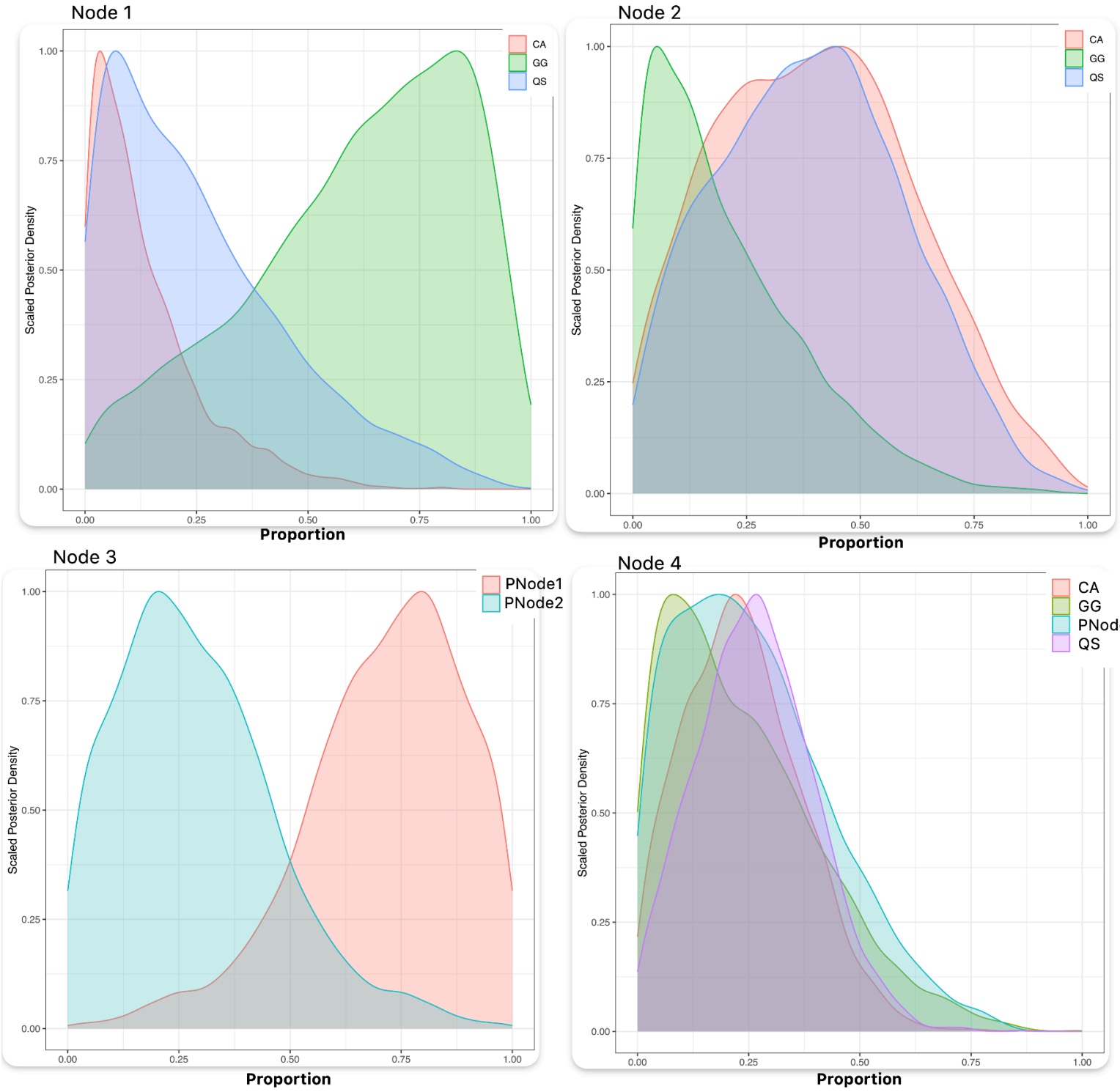

Figure 4.2 Posterior distributions of sediment source contributions at Nodes 1-4. This figure illustrates the scaled posterior density distributions of the upstream sources.

Node 1: The posterior distribution clearly indicates that GG is the dominant contributor, with the mode at 83%. The median and mean are also high, at 65 and 61%, respectively. However, the 95% confidence interval ranging from 7% to 95.2% reveals a high variability in the contribution from GG. In contrast, the contribution from CA is considerably less, with the mode at only 3%. The narrower 95% confidence interval for CA, from 0.4% to 46.6%, suggests less variability and greater certainty that the contribution from CA is lower at Node 1. QS has a moderate contribution with a mode at 6.9%. The 95% confidence interval for QS, from 0.8% to 75.8%, is less wide than GG's, indicating a moderate level of uncertainty in its contribution.

Node 2: The posterior distribution of sources indicates a different pattern of sediment contribution than what was observed for Node 1. Here, sources CA and QS emerge as the primary contributors, while GG shows a lesser influence. The mode for CA is at 46%, indicating that it has the strongest contribution of the three sources. Both the median and mean of CA is at 40% and the wide 95% confidence interval from 3.1% to 83.2% indicating significant variability. Conversely, GG presents a much smaller mode of 5%, suggesting lower contribution. The median and mean, at 16 and 21% respectively, further affirm GG's minor role in the sediment composition at Node 2. Its 95% confidence interval spans from 0.5% to 62.2%, reflecting a wide range of potential contributions but generally on the lower end. QS stands out as a substantial contributor, alongside CA, with a mode of 44%. Both the median and mean are around 39%, implying consistent contributions from this source. The standard deviation for QS, at 0.202, and a 95% confidence interval ranging from 3.7% to 78.9%, indicate high variability, yet with a narrower range than CA, suggesting slightly more certainty in QS's contributions at Node 2.

Node 3: The statistical summary and the posterior distribution graph combined gives a comprehensive understanding of the sediment proportion contribution from Node 1 and Node 2 to Node 3. The mode for contribution from Node 1 is 80%, which indicates that the contribution to the sediment at Node 3 is predominantly from Node 1. This high mode is

supported by both the median at 73 and the mean at 71%, suggesting a strong central tendency and reliability in the estimates. The standard deviation (SD) is relatively small at 0.180, indicating that the contributions are fairly consistent. Moreover, the 95% confidence interval, ranging from 26.9% to 98.2%, shows a wide range of possible values but with a high degree of certainty that at least 26.9% of the contributions come from P_{Node1} at the lowest end. The contribution from P_{Node2} mirrors the result from P_{Node1}, meaning that the contribution ratio between the nodes is either approximately 1:4 or 3:7. This is important for the interpretation of the final contribution of the full study site.

Node 4: The posterior distribution output from 'P_{Node3}', which consists of the combined contribution from Node 1 and Node 2, has a mode of 18%, a median of 24%, and a mean of 26%, suggesting that the most common value is slightly lower than the center of the distribution. The standard deviation of 0.172 indicates moderate variability in contributions. The percentile distribution shows a fairly narrow range, with the vast majority (97.5%) of contributions being below 65%. CA has a mode of 22%, a median of 23% and mean of 0.237. The standard deviation is 0.131. The contributions are generally consistent, with 97.5% of values not exceeding 66%. GG has a mode of 8%, considerably lower than the median (20%) and the mean (24%). The standard deviation of 0.174 signifies a moderate spread of data. Contributions from GG extend to a higher range, with 97.5% of values below 0.659. QS has a mode of 27%, which is the highest among the contributions at Node 4, and is very close to the median of 26% and the mean of 27%. This narrow clustering of the central tendency measures indicates a stable contribution. With a standard deviation of 0.130, the QS contributions exhibit less variability, as reflected in the 97.5th percentile value of 0.535, a narrower spread of values compared to other sources.

The result from the deconvoluted MixSIAR mixing model resulted in the posterior distribution graphs (Figure 4.3 and Table 4.3) for Node 3 and 4, illustrating the contribution from each source.

Table 4.3 Statistical Summary of the D-MixSIAR result.

	Source	Mode	Median	Mean	SD	2.5%	5%	95%	97.5%
Node 3	CA	0.142	0.190	0.211	0.126	0.033	0.047	0.451	0.512
	GG	0.527	0.504	0.495	0.201	0.103	0.145	0.807	0.846
	QS	0.214	0.270	0.294	0.165	0.044	0.066	0.615	0.673
Node 4	CA	0.282	0.287	0.292	0.128	0.066	0.093	0.511	0.553
	GG	0.313	0.351	0.366	0.167	0.087	0.117	0.663	0.731
	QS	0.336	0.340	0.341	0.132	0.093	0.125	0.560	0.603

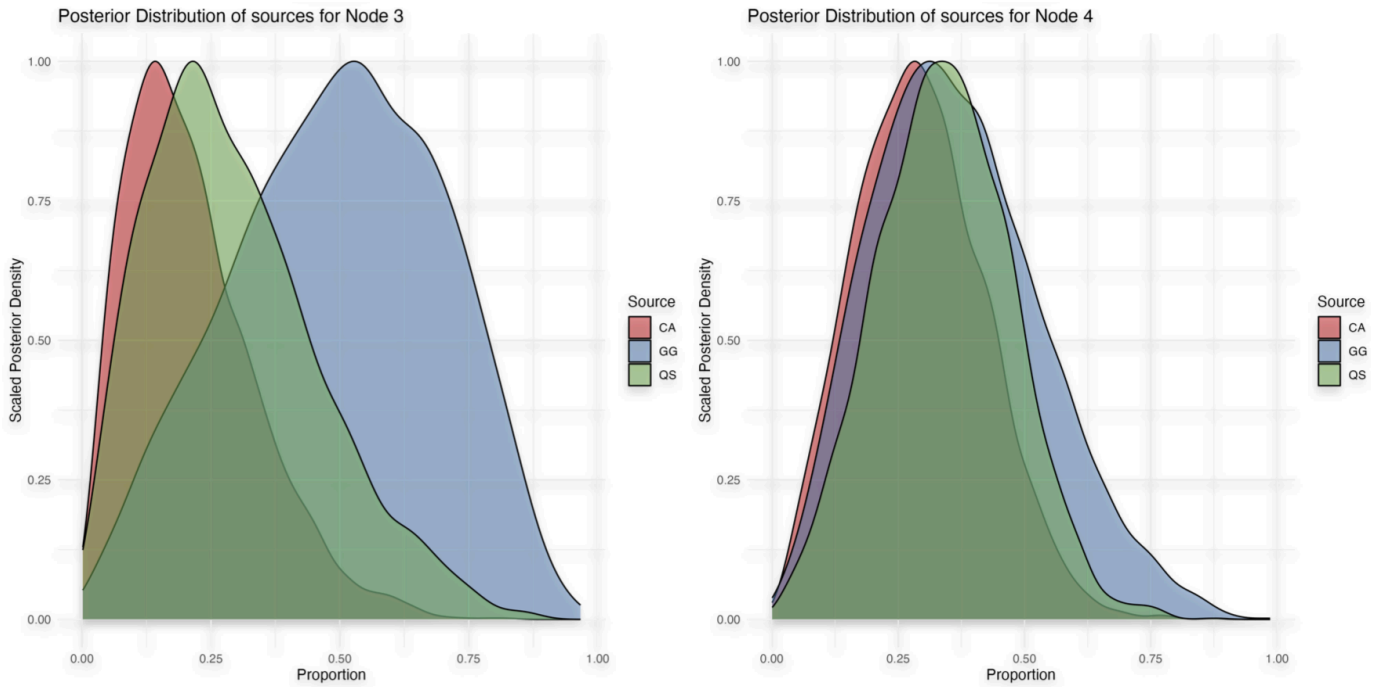

Figure 4.3 Posterior distributions of sediment source contributions at Node 3 and 4.

At Node 3, the deconvolution analysis in the D-MixSIAR framework integrates sediment contributions from upstream Nodes 1 and 2, offering insights into the compounded sediment dynamics. The posterior distribution for Node 3 indicates that GG is the leading contributor,

with a mode of 53%. This reflects the cumulative impact of upstream processes, where GG's prominence at Node 1 notably influences its continued dominance downstream. The median and mean, at 50%, confirm GG's substantial influence on the sediment profile at this juncture, which is supported by the relatively low standard deviation of 0.201. Conversely, the contributions from CA are less pronounced, with a mode of 14%. The median and mean values, 19 and 21% respectively, and confidence intervals ranging from 3.3% to 51.2%, suggest that CA, while less significant than GG, still contributes a notable portion of sediment to Node 3. QS shows a moderate level of influence with a mode of 21%, a median of 27%, and a mean of 29% – these figures position QS as a secondary but important contributor. The spread of QS contributions, with a standard deviation of 0.165 and a 95% confidence interval from 4.4% to 67.3%, points to a large variability that could be influenced by the differing source contributions from Nodes 1 and 2.

Finally at Node 4, the deconvolution analysis within the D-MixSIAR framework integrates sediment contributions from upstream Node 3, which itself integrates contributions from Nodes 1 and 2. Therefore, the results for Node 4 are not solely a reflection of local contributions but are also influenced by the upstream deconvolution process. The posterior distribution reveals a balanced mix of sediment contributions from the three sources, without a single source clearly dominating the composition. CA has a mode of 28%, a median of 28.7% and a mean of 29.2%, and the wide 95% confidence interval from 6.6% to 55.3% points to a significant degree of uncertainty. GG has a mode of 31%, median at 35.1% and mean at 36.6%, which implies a tendency for GG's contributions to be slightly higher than CA's. The confidence interval for GG, which stretches from 8.7% to 73.1%, reflects even greater variability than for CA. For QS, mode, median and mean are all at 34%, indicating a distribution that is relatively symmetrical around this central value. The confidence interval for QS, extending from 9.3% to 60.3%, denotes a moderate level of uncertainty in its contributions.

Turbidity analysis result

Turbidity measurements of streams from three campaigns are presented in figure 4.4, showing varying levels of turbidity in streams from smaller subcatchment. The turbidity is measured in Nephelometric Turbidity Units (NTU) and indicates how clear the water is; higher turbidity means more particles are present.

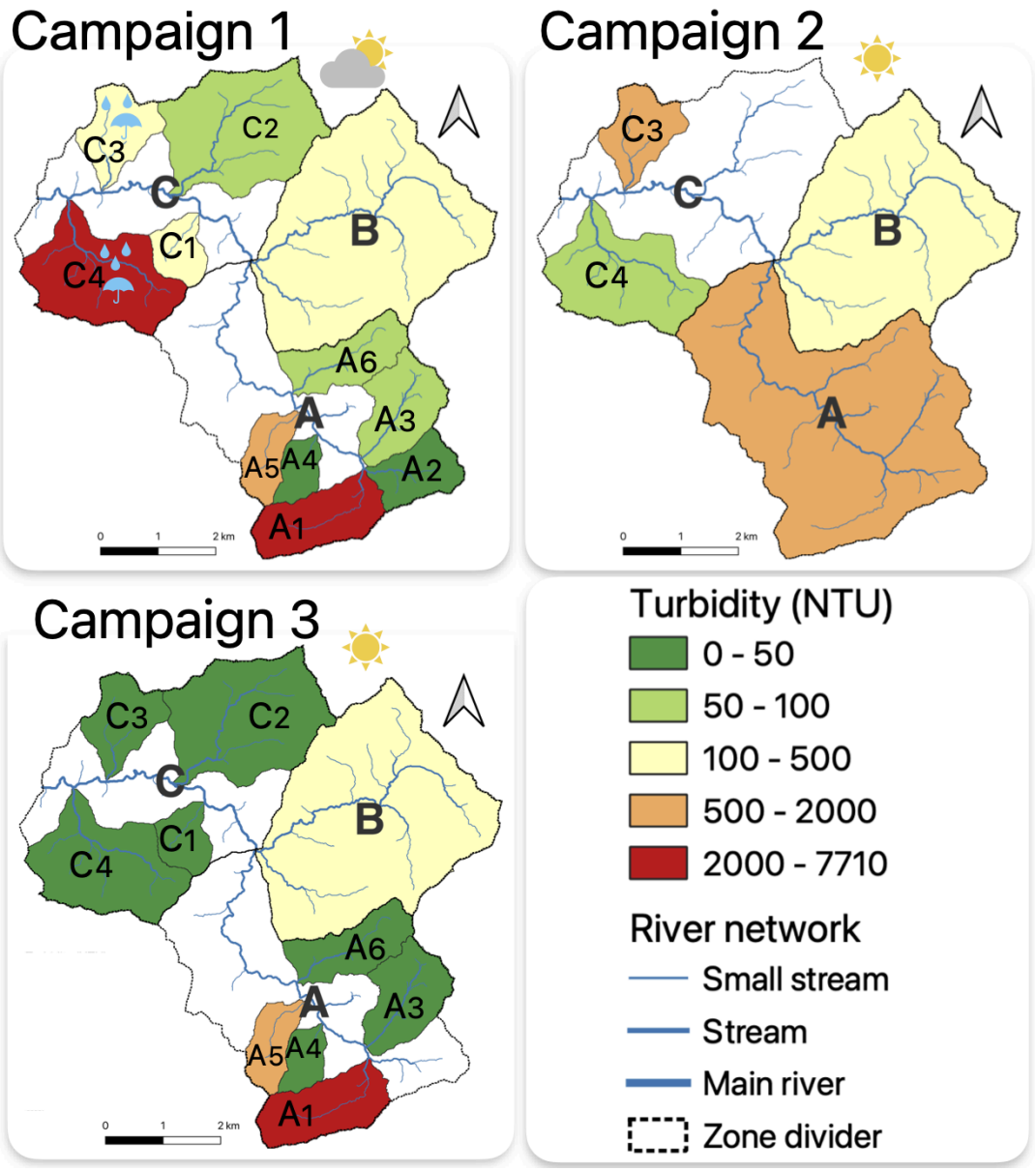

Figure 4.4 Turbidity measurements from streams. Created by author, 2023.

Campaign 1: Zone A shows ‘high’ (500-2000 NTU) and ‘very high’ turbidity (2000-7710 NTU) input from A1 and A5 respectively, indicating a high presence of suspended particles. The subareas A2 and A4 had ‘very low’ turbidity levels (0-50 NTU) and A3 and A6 had ‘low’ turbidity levels (100-500 NTU) respectively. Zone B had moderate turbidity. The measurements in C3 and C4 had ‘moderate’ (500-2000 NTU) to ‘very high’ turbidity respectively. During the measurements in Zone C, there were weather changes in these areas that led to high precipitation which probably increased the turbidity in these areas in the comparison to the measurements in the other zones.

Campaign 2: During this campaign, most of the samples were rejected, due to dirty containers, which increased the biotic activity and further altered the turbidity of the sample. The turbidity from the outlet of Zone A was ‘high’. Zone B's turbidity was ‘moderate’ as it was during the first campaign, which was expected, due to the similar conditions as during the first campaign. The turbidity measurements Zone C had a ‘high’ and ‘low’ contribution from C2 and C3 respectively. The high turbidity from C2 was not expected because the influence from heavy rainfall was not a factor, which indicates that other land disturbing activities are present from the area. C3 however had lower turbidity contribution compared to the first campaign, suggesting that rainfall events truly are an important factor on the turbidity.

Campaign 3: Zone A shows that A1 and A5 contain a ‘very high’ and ‘high’ turbidity respectively, which is similar to the first campaign. The other areas in Zone A were contributing with low turbidity. The contribution from Zone B was still moderate and the contribution from Zone C was low from all areas. The result during the final campaign indicates land disturbing activities.

The response of the stream input is illustrated with bar charts in figure 4.5, where each chart is the average turbidity at each sampling point.

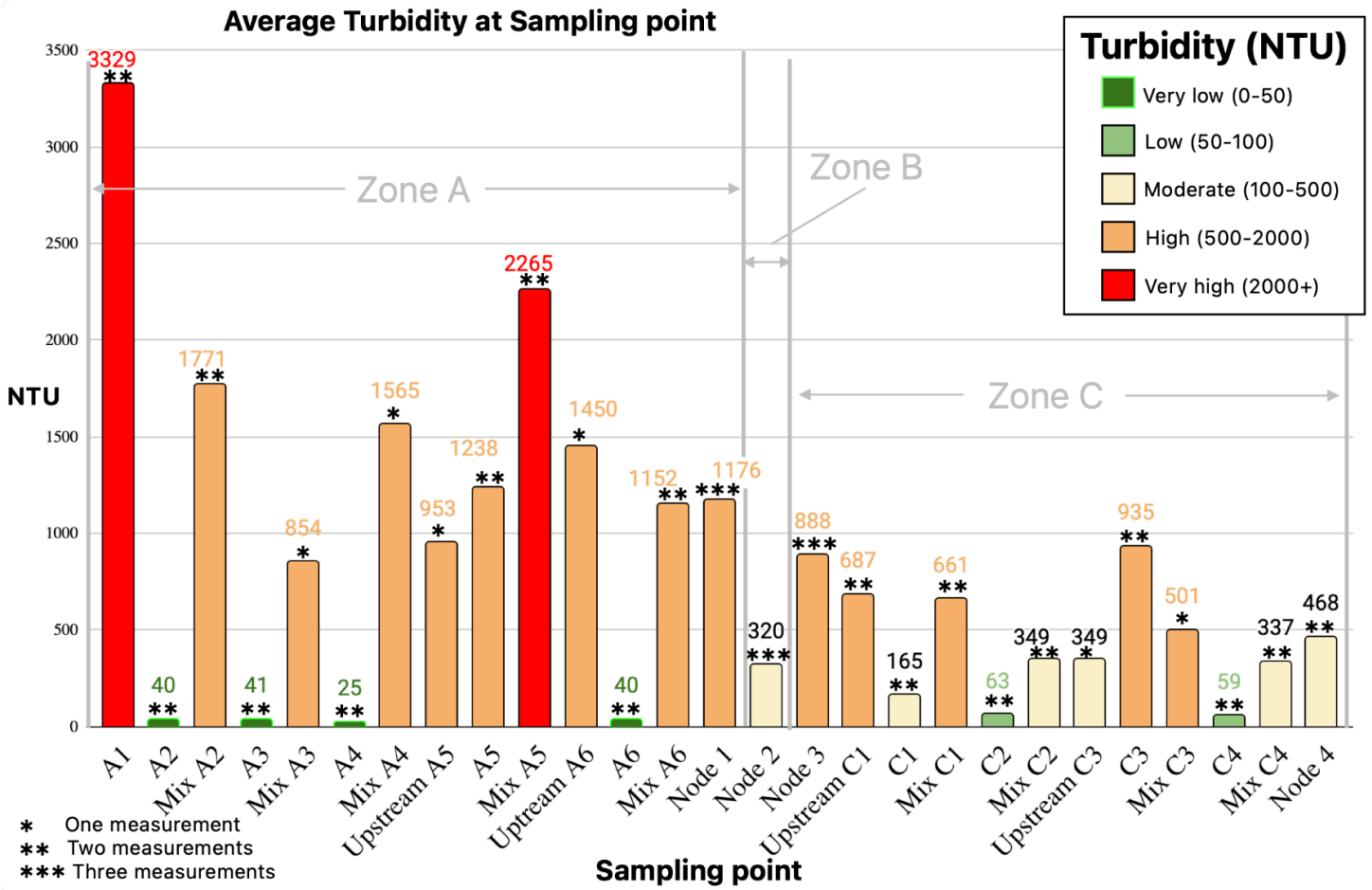

Figure 4.5 Average turbidity at sampling point. A1-6 represents streams contributing to the main river in Zone A, 'Upstream A5-6' is the turbidity just before the input of the stream, while 'Mix A2-6' represents the mixture after input. Same principle goes for the samples in Zone C. The amount of sampling occasions for each sampling point is denoted with '*'.

Zone B does not have any measured streams. Node 1-4 is the output from each zone, except Node 3 that represents the mixture of Node 1 and 2. All samples collected during the rainfall event were excluded because of its high influence noted in the previous section. Note that some of the sample points have only measurements from one or two occasions, while the maximum number of sample occasions are three. This result indicates that turbidity decreases, as more streams with low turbidity are mixed into the river. However, the average measurement of 'Mix A4' shows an increase in turbidity which is not expected because of the low turbidity input from A4.

Estimated turbidity contribution fractions from Node 1, Node 2, streams in Zone C, and upstream sources in Zone C are presented in table 4.4.

Table 4.4 Estimation of turbidity contribution from streams and upstream.

Location	Average Turbidity fraction (%)	Type
Node 1	66	*Upstream (Main river)*
Node 2	34	*Stream (Tributary)*
Node 3	**100**	*Downstream (Main river)*
Upstream C1	95	*Upstream (Main river)*
C1	5	*Stream (Tributary)*
Mix C1	**100**	*Downstream (Main river)*
Mix C1	48	*Upstream (Main river)*
C2	52	*Stream (Tributary)*
Mix C2	**100**	*Downstream (Main river)*
Upstream C3	74	*Upstream (Main river)*
C3	26	*Stream (Tributary)*
Mix C3	**100**	*Downstream (Main river)*
Mix C3	63	*Upstream (Main river)*
C4	37	*Stream (Tributary)*
Mix C4	**100**	*Downstream (Main river)*

Estimation of the turbidity contribution fractions from each zone to the final output at Node 4 is presented in table 4.5.

Table 4.5 Result of the turbidity contribution fractions and NTU.

Zone	Turbidity Contribution fraction (%)	NTU
Zone A	14.0	66
Zone B	7.1	33
Zone A+B	**21.1**	**99**
Zone C	78.9	369
Node 4	**100**	**468**

MixSIAR and turbidity comparison

A comparison between the estimated contribution fractions based on the turbidity measurements and the fingerprinting model outputs is presented in figure 4.6.

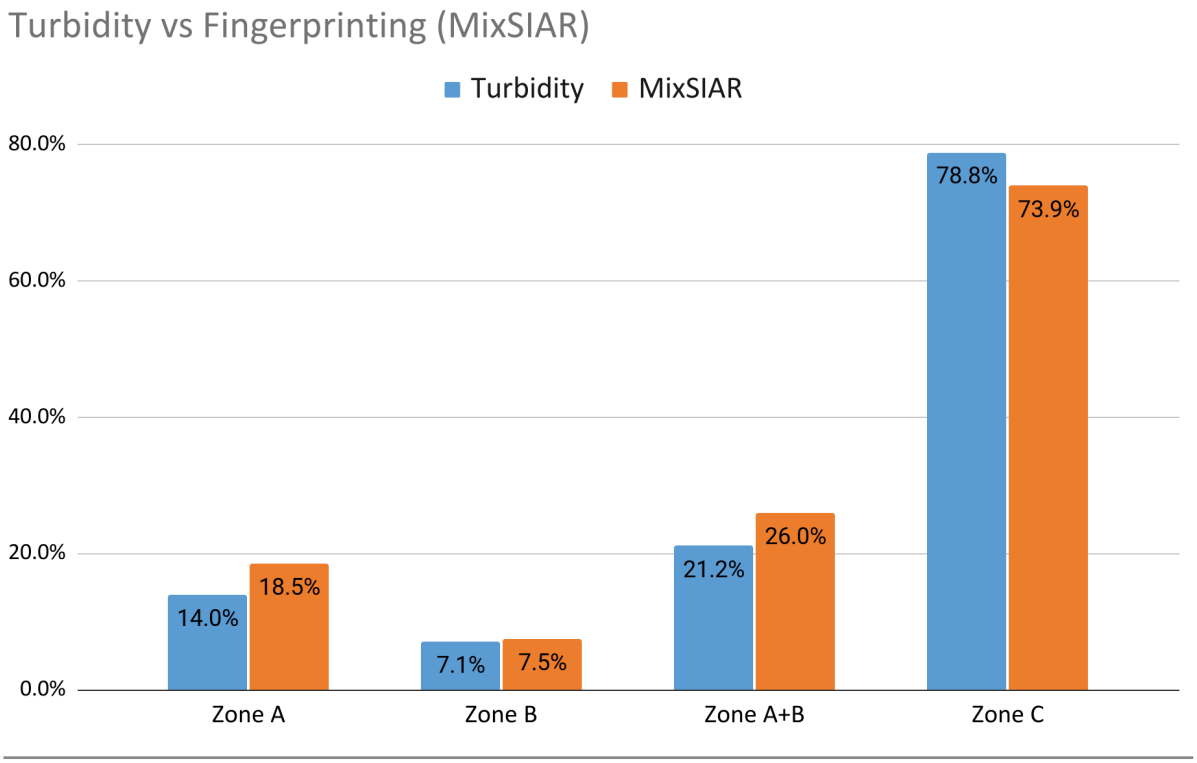

Figure 4.6 Comparison of the estimated contribution fractions from the turbidity computations and MixSAIR models.

The difference between the contribution fraction result at nodes from the turbidity analysis and MixSIAR modeling result is 4.5% in Zone A, 0.4% in Zone B, 4.8% in Zone A+B and 4.9% in Zone C. This result indicates that there is a correlation between the turbidity contribution and the source contribution. The similarity in the result between the two independent methods also serves to validate the results – indicating that both methods give reliable estimates of the relative source contributions.

MixSIAR and RUSLE comparison

The land degradation map over Rwanda created by PKabatha (2023), based on the Revised Universal Soil Loss Equation (RUSLE), was employed to create a map of land degradation potential in the study site, see figure 4.7. The associated classification coverage in hectares for each zone are presented in table 4.6.

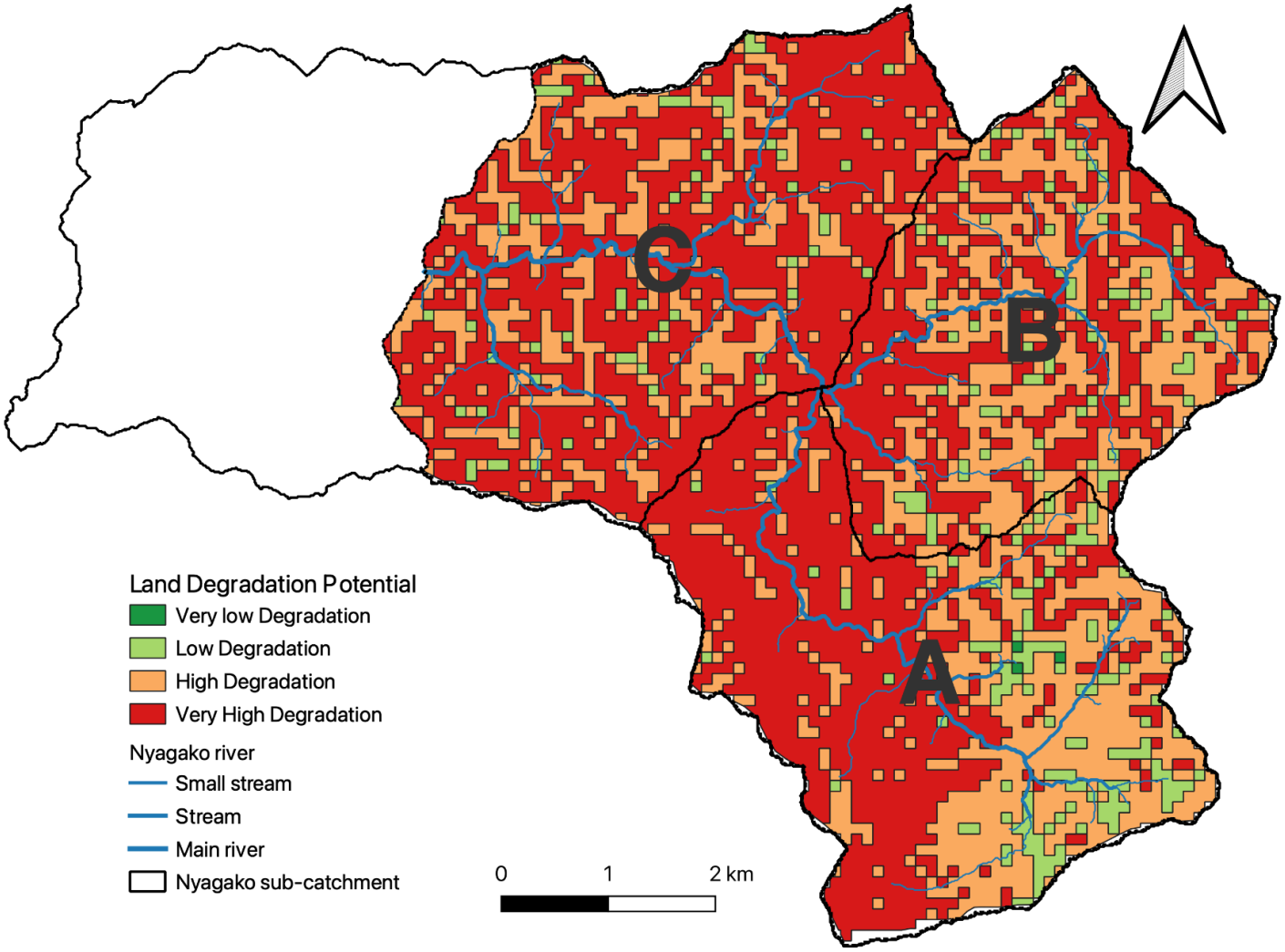

Figure 4.7 Estimated land degradation of the study site within Nyagako sub-catchment, developed using the Revised Universal Soil Loss Equation (RUSLE). Created by author, 2023 (Source: PKabatha, 2023).

Table 4.6 Land degradation potential distribution in hectare and percentage for each zone of the study site.

Land Degradation Potential	Zone A	Zone B	Zone C
Very Low Degradation:	3 ha (0.2%)	–	–
Low Degradation:	126 ha (8%)	77.6 ha (7%)	63.6 ha (5%)
High Degradation:	566.7 ha (38%)	519.7 ha (46%)	478.3 ha (35%)
Very High Degradation:	809.7 ha (54%)	546.7 ha (49%)	1046.1 ha (75%)

The land degradation assessment reveals significant soil erosion issues, especially in Zone A and C. All three zones predominantly exhibit high to very high land degradation. Zone C has the largest fraction of 'Very High Degradation', corresponding to 75% of the total area in Zone C. Zone A and Zone B have a similar fraction of 'Very High Degradation', 54 and 49%, respectively. This suggests that Zone C is the most erosion prone area according to RUSLE, similar to the MixSIAR and turbidity result, although there is not a direct comparens as between MixSIAR and the turbidity.

4.2. Potential Drivers of Soil Erosion

LULC combined with lithology and zones

Three stacked bar graphs for each zone in figure 4.8 illustrate the lithology type distribution including the LULC distribution.

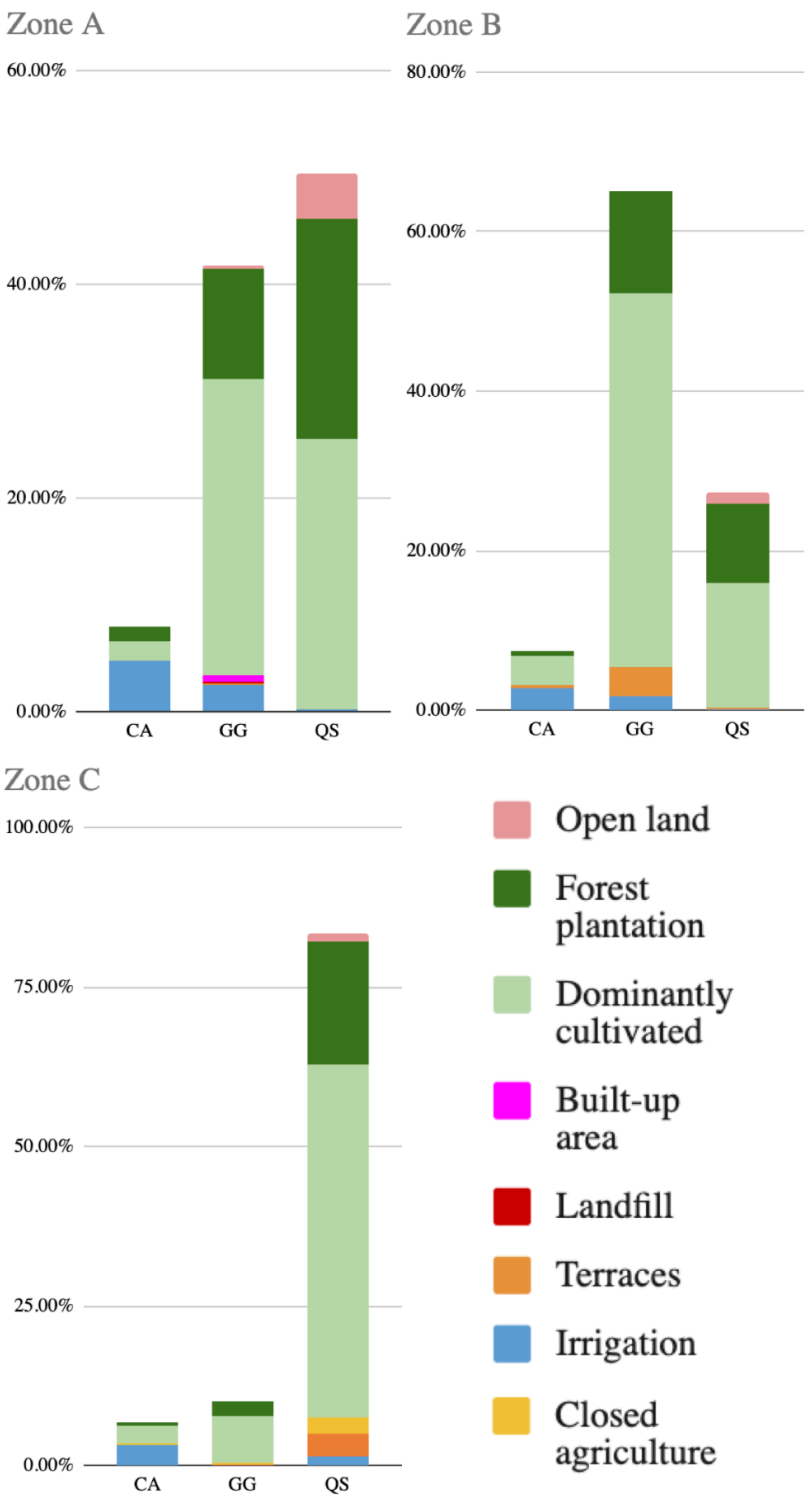

Figure 4.8 Land use distribution across the lithology types CA, GG and QS for each zone: A-C.

The distribution of the lithology types across the zones varies, except that CA is consistently lowest. For Zone A, QS is slightly more distributed than GG, while dominated in Zone C, where GG and CA are almost evenly distributed. GG is the dominating lithology type in Zone B. Dominantly cultivated is a major land use in all zones across the lithology types. Although, irrigation is the most dominant source of CA. Forest plantation land is the second most prevalent land use. Other land use types like closed agriculture, terraces, irrigation, open land, built-up areas, and landfill represent smaller proportions of land use and are not consistently present across all zones. While there is some variation in land use between lithology types, especially for minor land use types. The distribution patterns are consistent across the zones.

LULC distribution for each source across the different zones A-C, categorized into Forest, Cultivated, and Non-cultivated, is depicted in Figure 4.9.

Figure 4.9 Land use distribution for each source (CA, GG and QS) across the zones A-C.

This result illustrates the cultivation-forest ratio for each source across the zones. The cultivated land use contribution is consistently lower in 'Zone A' among all of the sources and increases gradually for each zone, except for Zone C in GG. Both forest and non-cultivated land use is dominating in the source of QS, with a gradual decrease from zone A to C. Forest cover is generally higher in Zone A, among the sources.

Slope distribution

The slope distribution is delineated into four categories based on their slope gradient. The intricacies of these slope classifications are visually captured in figure 4.10, which presents a detailed map of the slope distribution across the study site.

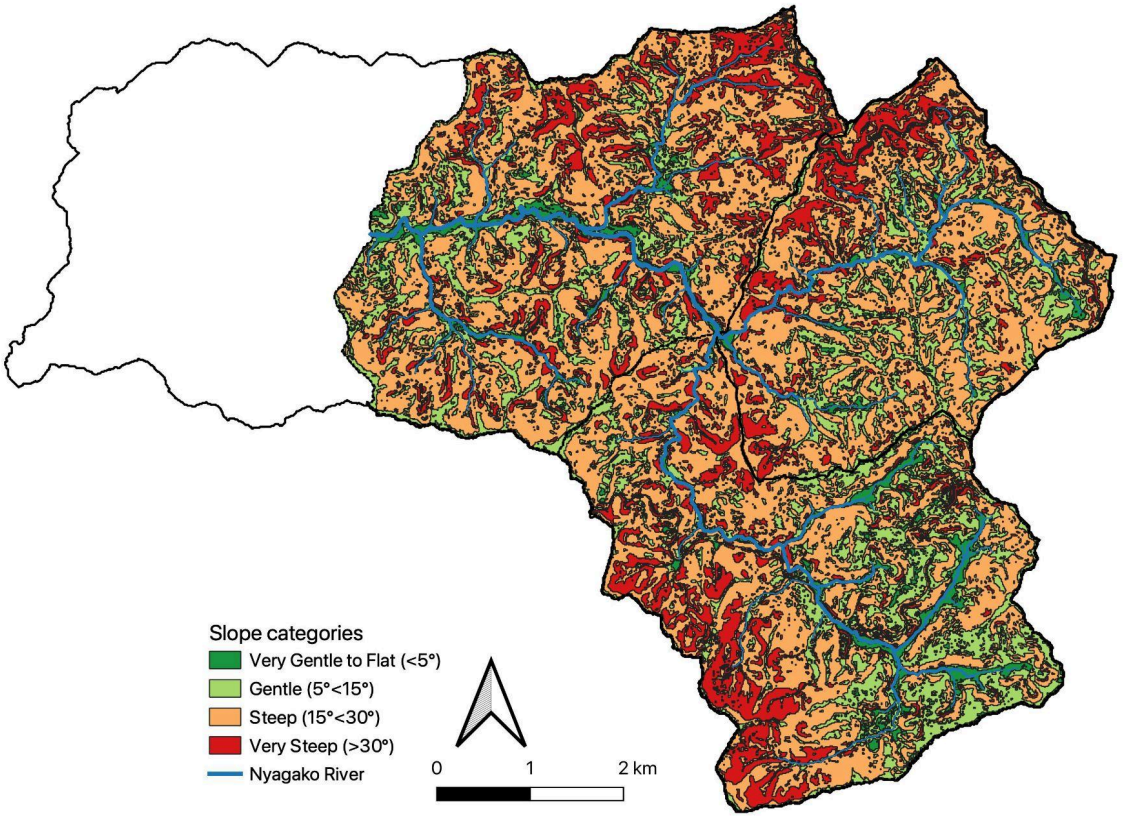

Figure 4.10 Slope distribution of the slope categories: 'Very Gentle to Flat', 'Gentle', 'Steep' and 'Very Steep'. Created by author, 2023 (Source: University of Rwanda).

The majority of the catchment, accounting for 54.7%, falls into the 'Steep' slope category with gradients ranging from 15° to 30°. These areas are particularly susceptible to rapid surface runoff and may pose significant challenges for land management practices. In contrast, 24.4% of the catchment features 'Gentle' slopes, with angles between 5° to 15°, potentially favoring agricultural activities due to lower erosion risks and easier cultivability. The 'Very Steep' slopes, characterized by angles exceeding 30°, comprise 15.8% of the site. This category is likely to experience the highest rates of soil erosion and may also present natural habitats less disturbed by human activity due to the difficulty of access and utilization. Remarkably, only a small fraction, 5.1%, is characterized as 'Very Gentle to Flat' slopes with angles less than 5° – these areas may serve as zones of sediment deposition because they are located in the lowest part of the valley and are exposed to flooding and channel collapse during high flow.

Comparison with modeling result

Table 4.7 summarizes the results used for the analysis of causes of soil erosion. It presents the contribution fractions for each source and zone, size of area, corresponding contribution per area, average slope gradient, and cultivated-forest ratio. Table 4.7 is presented as a complement for interpreting the result in the following sections.

Table 4.7 Summarized results of the contribution fractions and potential causes of soil erosion.

	Source	Final Contribution*	Contribution in zone	Area (ha)	Contribution in zone /ha**	Avg. slope gradient	Cultivation-forest ratio
Zone A	CA	2.5%	13.5 ± 12.6	121	0.11%	12°	4.39
	GG	11.3%	60.9 ± 24.2	631	0.10%	17°	2.90
	QS	4.7%	25.6 ± 20.2	765	0.03%	23°	1.24
		18.5%	***100%***	***1516***		***19.8°***	
Zone B	CA	3.0%	40.3 ± 21.5	86	0.47%	14°	9.15
	GG	1.5%	20.5 ± 16.9	737	0.03%	18°	3.82
	QS	2.9%	39.2 ± 20.2	325	0.12%	25°	1.59
		7.5%	***100%***	***1148***		***20.2°***	
Zone C	CA	23.7%	32.1 ± 12.6	111	0.29%	13°	12.68
	GG	23.7%	32.1 ± 20.1	163	0.20%	16°	3.61
	QS	26.5%	35.9 ± 16.5	1316	0.03%	22°	3.05
		73.9%	***100%***	***1589***		***21.1°***	
All zones	CA	29.2%	-	318.3	0.09%	12°	8.74
	GG	36.6%	-	1530.7	0.02%	18°	3.44
	QS	34.1%	-	2404.6	0.01%	23°	1.96
		100%		***4253.6***		***18°***	

* Final contribution fraction per source and zone (+ 'All zones') at Node 4.
** Note, final contribution are used for 'All zones'.

The relationship between the contribution per hectare and the average slope angle is scatter plotted in figure 4.11.

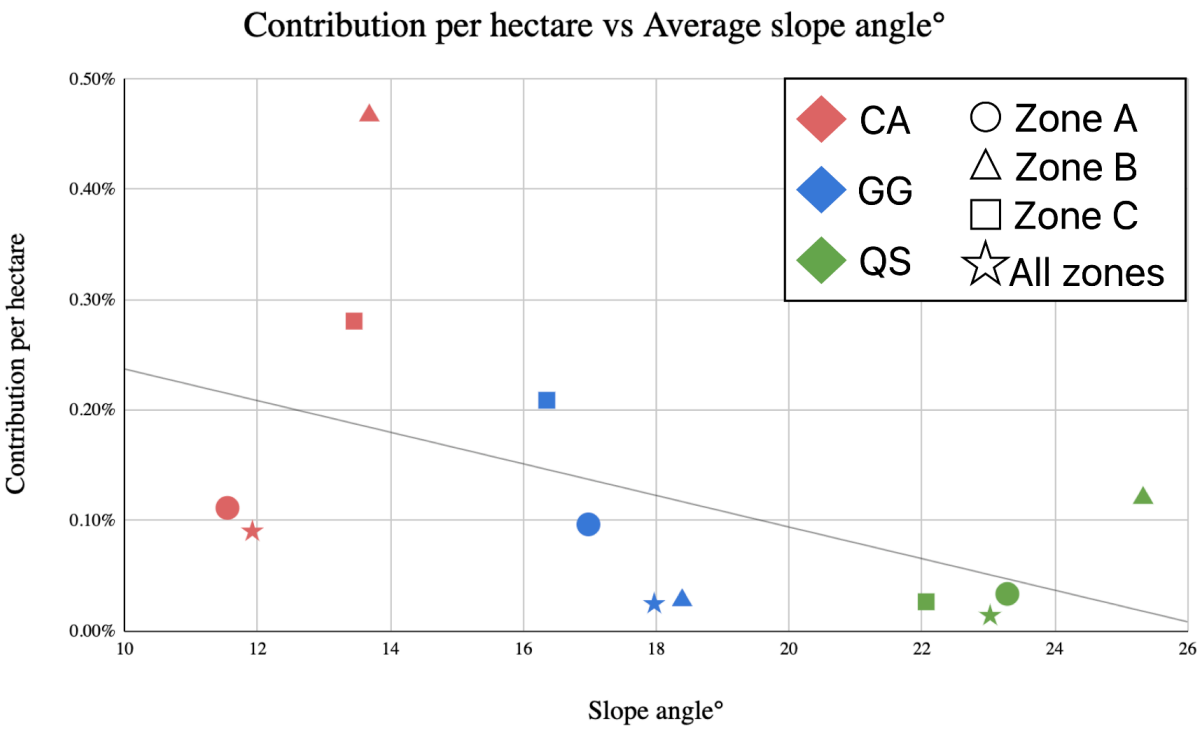

Figure 4.11 Scatter plot of contribution per hectare against the average slope angle for each zone and for the full study site (All zones).

An inverse correlation is observed, with an increasing slope angle corresponding to a decreasing contribution per hectare. Source-specific slope contributions are as follows: CA predominantly falls within an 11-14° range, GG within 16-19°, and QS within 22-26°. Notably, the contribution in Zone B for CA is substantially higher (0.47%) compared to the other zones.

The scatter plot in figure 4.12 illustrates the correlation between soil erosion contribution per hectare and the cultivation-forest ratio, with sources differentiated by color and zones by symbols:

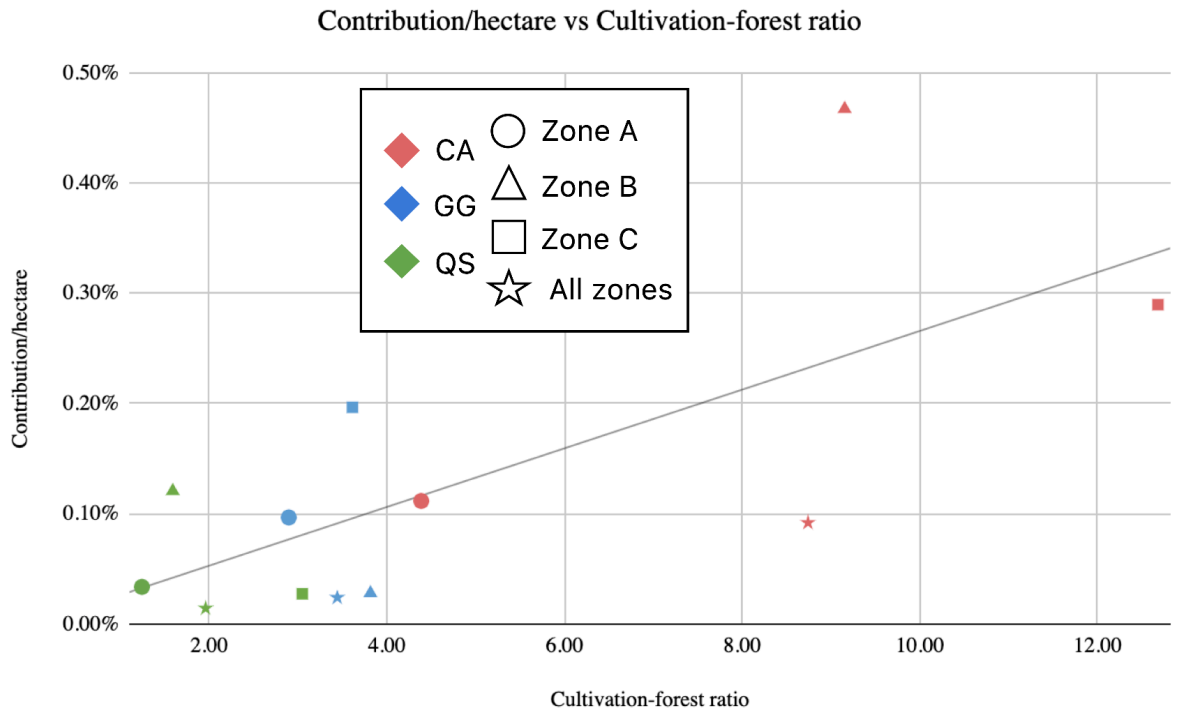

Figure 4.12 A detailed correlation scatter plot highlighting the relationship between per-hectare soil erosion contribution and cultivation-forest ratios, segregated by the sources CA, GG and QS and zones.

There is a trend wherein areas with a higher cultivation-forest ratio tend to exhibit an increased contribution of soil erosion. The distribution of the sources indicate that the cultivation-forest ratio is generally higher for CA, then GG and lowest for QS.

Additional drivers from field observations

Mining activities, both legal and illegal, were observed in all three zones. The most significant visual as was noted in the upstream regions of Zone A, see figure 4.13. Mining sites were frequently directly within the main river or adjacent to streams. Resources such as coltan, wolfram, or sand for construction were among those extracted from these sites.

Figure 4.13 Impact from mining activity in Nyagako catchment. Top pictures: Mining site in Zone C. Bottom left: Junction between outlet from A1 and A2 in Zone A. Bottom right: Junction between A3 and upstream mix of A1 and A2 in Zone A.

During the first sediment campaign, a heavy rainfall event occurred, altering the run off, see figure 4.14. Rill erosion and runoff within gullies started to occur. This had a clear impact on the turbidity and possibly the sediment sample collected at node 4.

Figure 4.14 Impact of rainfall in Nyagako catchment. Top: Increasing run off from agricultural land in Zone C. Bottom: The left picture is influenced by rainfall, the right picture is the same location (C4 in Zone C) with no recent influence by rainfall event. The red dots indicate the location of the same rock in the river for comparison.

5. Discussion

5.1. Erosion-Prone Zones and Sources

The highest contribution per hectare was from CA in Zone B, corresponding to a contribution of 0.47% per hectare. The second highest was CA in Zone C (0.29%), followed by GG in Zone C (0.20%). However, according to the final contribution per hectare at Node 4, the result revealed that CA in Zone C has the highest contribution per hectare, corresponding to a contribution of 0.21% per hectare, followed by GG in Zone C (0.15%). Overall, CA showed the highest contribution per hectare across all zones, which suggests that areas with CA have the highest erosion rate.

Both the sediment fingerprinting and turbidity analysis estimated that Zone C had the highest contribution to Node 4, estimated to 78.8% and 73.9% respectively, followed by Zone A, at 14.0% and 18.5%, and Zone B, with 7.1% and 7.5%. In addition, according to the land degradation map provided by PKabatha in 2023, Zone C has the highest potential for land degradation among the zones, with 1046.1 hectares classified as 'Very High Degradation', corresponding to 75% of the area in Zone C, and 478.3 hectares as 'High Degradation' (35%). In comparison, Zone A has 809.7 hectares classified as 'Very High Degradation' (54%), and 566.7 hectares as 'High Degradation' (35%), indicating it has the second highest land degradation potential. Zone B has 546.7 hectares classified as 'Very High Degradation' (49%) and 519.7 hectares as 'High Degradation' (46%). This aligns with zone-based contribution estimation from both the sediment fingerprinting and turbidity analysis.

According to the mean value of the posterior proportion output from the D-MixSIAR model result, the final contribution fraction from the different lithology types (i.e. the output from Node 4) was relatively even distributed: CA at 29.2 ± 12.8% , GG at 36.6 ± 16.7%, and QS at 34.1 ± 13.2%. It is important to note that GG has a mode of 31%, while QS has a mode of 37%, this implies that QS's most frequent contribution fraction is higher than GG's. However

there is not an obvious difference, which implies that physical and/or chemical, anthropogenic factors interplay and equalize the contribution among the sources.

The mean values of the posterior proportion output from Node 3 – CA (21.1 ± 12.6%) , GG (49.5 ± 20.1%) and QS (29.4 ± 16.5%) – suggest that GG is the most significant contributing upstream source. Both the mode (52.7%) and median (50.4%) are close to the mean value, which indicates an even distribution graph. Although, there is relatively high standard deviation, which adds some uncertainty. However, this result suggests that GG within Zone C should have a low contribution. The MixSIAR result, which estimates the contribution from each zone, confirms that suspicion, especially by the mode value for GG (8%), which represents the most frequently posterior contribution value.

Tracing the contribution of GG further upstream (Zone A and Zone B), there are strong indications that the major contribution originates from Zone A, estimated to correspond to 11.3% of the final contribution at Node 4. The contribution of GG in Zone A has a mode of 83.2% and a mean value (60.9 ± 24.2%). The mean posterior distribution output for the other sources in Zone A; CA and QS was (13.5 ± 12.6%) and (25.6 ± 20.2%) respectively.

The study by GLOWS-FIU (2016), made different choices regarding source selection compared to this study. The decision not to replicate the same set of sources stemmed from the observation during PCA iteration, which revealed a more distinct source distribution and affiliation among CA, GG, and QS in this study. However, despite these disparities, significant similarities and correlations persist between the source sets, warranting some comparison. The study by GLOWS-FIU identified that 'Bb/Ng' (Bumbugo/Nyabugogo formations) emerged as the dominant contributor for the catchment, constituting approximately 36-60% of the total contribution. This source primarily comprises quartzite and sandstone. In this study, this source distribution aligns with the source group QS, which accounts for approximately 34.1% of the total contribution. Similarly, ‘Gd’ (Granitoides divers) contributing less than 22% and 'Gdm' (Granites to Mica) contributed approximately

9-29%, which in this study collectively correspond to the GG source group, contributing approximately 36.6%. However, a notable departure lies in the inclusion of CA as a source group in this study, corresponding to the geological type 'Ho' in the study by GLOWS-FIU (2016). It is worth noting that the distribution of 'Ho' was not as widespread as CA. ‘Ho’ was omitted as a source due to its association with alluvial deposits, which were perceived as less advantageous for accurate source identification due to their complex mixture of upstream origins.

Further important disparities between these studies are that this study does not cover the full extent of Nyagako sub-catchment and is not conducted during the same temporal periods and coverage. The sampling period for this study was from September to December, which also included the shorter rainy season (October to November), while the GLOWS-FIU study was conducted from the end of January to the end of April, when the longer rainy season occurs (March to May). Additionally, methodological distinctions are apparent, with this study employing a deconvoluted MixSIAR mixing model approach, as opposed to the GLOWS-FIU study's pooling of all soil samples from the catchment against sediment samples solely at the catchment's outlet. Therefore, a direct comparison of the results would not be adequate for any direct conclusion or validation of the previous result, although the comparison of the results is valuable for future similar studies.

5.2. Contributing Factors

The analysis of the contributing factors highlighted a significant role of LULC distribution. Given the crucial role of vegetation, especially forest cover, it was decided to simplify the classification of LULC into three main categories: forest, cultivation, and non-cultivated land. This approach aimed to uncover potential correlations between soil erosion and the ratio of cultivation to forested areas.

The results revealed a positive trend, indicating that higher ratios of cultivated to forest areas were associated with greater contributions per hectare. This trend suggests that regions with a higher intensity of agricultural activities relative to forest cover are more prone to soil erosion. This finding aligns with existing literature, which emphasizes the pivotal role of vegetation in mitigating soil erosion. As noted by Morgan (2005), plant cover effectively dissipates the energy of flowing water by introducing roughness to the flow, thereby reducing its velocity. The degree of roughness varies depending on factors such as plant morphology, density, and height in relation to the flow depth. Additionally, the GLOWS-FIU study highlighted that open agriculture plays a significant role in sediment levels due to its extensive coverage, and the combination of these activities with localized rainfall events can further intensify sediment contribution.

Another observed trend was that the lithology sources – QS, GG, and CA – had an increasing trend of cultivation-forest ratio in consecutive order (i.e., QS has a lower cultivation ratio than GG, and GG has a lower ratio than QS). The explanation for this might be connected to the average slope gradient. Steepness and slope length are known to have a strong influence on soil erosion (FAO, 2019). The analysis of the topography of the study site revealed a highly hilly landscape, with elevations varying considerably from 1474 to 2184 meters above sea level. The average slope gradient for each source and zone was estimated, revealing the features among the sources: QS had an overall slope gradient of 23°, GG 18°, and CA 12°. This suggests that QS are generally located on higher elevations with steeper slopes, while CA are located deeper down in the valleys with lower slopes, and GG are situated somewhere in between. The connection to the positive contribution trend of the lithology sources in the order QS, GG, and CA regarding the cultivation-forest ratio could be that it is more difficult to cultivate on steeper slopes, fertile soil often already has been eroded and deposited deeper in the valleys where the slope gradient is lower. Therefore, forest plantations are more suitable for the steeper regions to mitigate further erosion.

GLOWS-FIU (2016) found that mining activities greatly amplify sediment production, especially during specific rainfall events at the site. In this study mining activities were observed to have a significant effect on the sediment contribution, evident from the turbidity measurements, particularly in sub-catchments A1 and A5 in Zone A. An interesting observation was a reduction in turbidity during the third campaign, which could be indicating less mining activity since it was a Sunday, which is typically associated with reduced work activity.

The impact of heavy rainfall events on soil erosion was observed, particularly evident in turbidity measurements during the comparison of each campaign for sub-catchment C3-C4 in Zone C. However, despite C3 experiencing significant rainfall during the first campaign, it exhibited higher turbidity levels during the second campaign, even in the absence of rainfall. Several factors could explain this discrepancy. Firstly, the difference in rainfall duration between C3 and C4 during the turbidity measurements could have influenced the results. Secondly, due to time constraints and permission to use the turbidimeter in situ, the turbidity measurements were conducted only once or twice at each sampling point – the first campaign in situ and the second and third campaign off-site, at the soil laboratory. Conducting the turbidity measurements off-site during the second and third campaigns may have made the samples more susceptible to biotic activity, resulting in the cancellation of many samples. Additionally, it is possible that there were human error issues with sample collection during the second campaign, leading to inaccuracies in the turbidity measurements. Ideally, conducting at least three in situ measurements at each location would have provided more accurate results. However, the study was conducted during dry conditions to ensure consistency across campaigns, albeit heavy rainfall events are known to be the primary driver of soil erosion. Despite this limitation, the findings shed light on the potential impact of rainfall variability on soil erosion dynamics.

GG in Zone A

Upon further investigation to explain the prominent contribution of GG in Zone A, an analysis of correlating factors such as the cultivation-forest ratio and average slope gradient among the sources did not resolve any significant correlation to the general trends, suggesting that other factors may be influencing soil erosion dynamics.

The cultivation-forest ratio for each source in Zone A was estimated to be 4.39 for CA, 2.90 for GG, and 1.24 for QS, pointing towards a higher expected contribution from CA. Similarly, considering the average slope gradient for each source; CA (12°), GG (17°) and QS (23°), according to the general trend, one would also anticipate the highest contribution from CA.

Furthermore, the turbidity measurements indicated a significant input of turbidity in area 'A1' as well as from 'A5,' which implies substantial land disturbance within these locales. The LULC map identified 10 mining activities in area 'A1' and 1 in area 'A5'. However, the location of these mining sites does not correspond with the distribution of GG and rather correspond to the distribution of QS. Either new or unrecorded (illegal) mining sites could be occurring within the area of ‘A1’ and ’A5’ (where GG are distributed), could potentially explain the high contribution of GG. Another potential source that could influence the result is an observed landfill in Zone A.

CA as the primary contributor per hectare

The discrepancy in distribution among the sediment sources, with CA emerging as the primary contributor despite its lower coverage compared to QS and GG, raises important questions about the validity of the fingerprinting results. QS covers the largest area, 2404.6 hectares, which represents 56.5% of the total area, followed by GG, 1530.7 hectares (36%), while CA accounts for a smaller portion of the study area, 318.3 hectares (7.5%). This incongruity prompts a closer examination of the factors influencing sediment contribution to ensure the accuracy and reliability of the analysis.

CA encompasses both alluvium and colluvium, representing sediment deposits transported and deposited by water and mass-gravity movement, respectively. Alluvium typically accumulates in channelized watercourse systems like gullies, streams, and floodplains, leading to well-sorted and stratified deposits. On the other hand, colluvium consists of sediments moved primarily by gravity, often found on slopes and their bases (Miller & Juilleret, 2020). Given these characteristics, CA is inherently exposed and susceptible to erosion, which could explain why its contribution per hectare exceeds that of other sources.

Zone C as the major contribution region

As previously established, Zone C contributes significantly to the final sediment output, accounting for approximately 70-80% of the total. Upon closer examination of the contribution fractions from Zone C, an analysis of the modes reveals a distinct pattern. The mode contribution of QS is the highest, about 26.6%, with the GG following 22.1% and 8% of GG, 18.2% originating from the upstream zones. The coverage of QS within Zone C is almost 12 times larger than CA and more than 8 times larger than GG in Zone C. This substantial difference underscores QS as the most significant contributing source at the Zone C outlet, suggesting that its actual contribution may even exceed the estimate. However, the average slope gradient is relatively large 22°, compared to the other sources 16° and 13° for GG and CA, respectively. As indicated by the general trend, a greater slope gradient correlates with a lower cultivation-forest ratio and, consequently, a lower contribution.

Overall, Zone C exhibits the highest cultivation-forest ratio among all sources, except for GG in Zone B, however being very close to the same ratio in Zone C. Additionally, Zone C has the highest average slope gradient, this might be an indication that extensive cultivation on steep slopes might trigger alluvial and colluvial erosion processes.

5.3. Evaluation of the Methodology

The methodology provided valuable insights for future investigations addressing soil erosion for different points of views, and the combination of the methods complemented the interpretation of the result.

The GIS-analysis was an essential tool, from the preparatory investigation for the fieldwork and project design and to the interpreting of the result. The existing land degradation map, based on RUSLE, provided support and insights. Applying a RUSLE model has a lot of potential to provide insights and understanding of the soil erosion dynamics. It is a relatively simple and widely approved model (Zeghmar et al., 2024). The limitation by utilizing GIS is that it relies on data points and layers, which is important to validate and confirm. This suggests that more investments on data collection correlating to the factors employed in the RUSLE model would be valuable.

The turbidity analysis generated valuable information that complemented and validated the fingerprinting results. It was furthermore an efficient method to map and back track smaller sub-catchments of interest and indications of anthropogenic land disturbance, like mining activity and highlighted the impact from rainfall events.

The sediment fingerprinting analysis successfully generated posterior density plots illustrating the distribution of contributions from the selected sources. The standard deviations (SD) and percentile ranges provide a sense of the variability and uncertainty in these estimates. The variability in source contributions underscores the complexity of sediment transport and deposition in the catchment and the influence of upstream activities and natural processes. The accuracy of the result is difficult to validate and confirm. However, the correlations to the turbidity analysis and land degradation map validate the result regarding the zone contribution. Furthermore, the comprehensive tracers screening process assures the quality of the data underlying the results. The fingerprinting method, compared to the RUSLE model and turbidity analysis, estimates what the sediment in the river contains and its lithological

origin, while the turbidity analysis can identify the direction of the origin of the soil erosion and RUSLE can identify the contributing areas.

Regarding the sampling strategy

This study was mainly focusing on collecting subsoil, because it has less content of organic matter and humus, although topsoil was often collected as a complement. Both layers have particle content of sand, silt and clay. The topsoil is more exposed to anthropogenic activity, which could contaminate the lithogenic source (Kříbek er al., 2010).

The goal of the soil collection was to obtain the geochemical signature of the lithology at the study site. During the analysis of the samples for the tracer screening process, the distinction between topsoil and subsoil was always considered. However, there was no obvious discrepancy between their geochemical signatures. The low variation in the composition of geochemical elements between topsoil and subsoil could be due to limitations in soil formation. Additionally, the confining deposits from alluvial and colluvial sources further limit soil development (Smith et al., 2018).

More soil samples from the zones would have been preferred. For instance, no soil samples of CA were collected in Zone B, so the average result for CA representing the whole study site was used to complement the MixSIAR modeling for Zone B. Similarly, the average result for QS was used in Zone B, and the average result for GG was used in Zone C due to a lack of soil samples representing those sources and zones.

Due to limited access, budget, and experience, it was difficult to reach and collect samples from all locations of interest. Furthermore, the final selection of sources was not determined prior to the soil sample collection campaigns; instead, the process was based on trials with different combinations using the geographical information available, which resulted in a lack of soil samples. Also, soil samples that were very close to a source boundary and had a strong

correlation to the geochemical signature of the neighboring source, were further on considered as part of that neighboring source.

The sample preparation was not performed in the most clinical environment, where several factors could influence and contaminate the result. The equipment, such as sieves, container, and plates were cleaned with distilled water before usage. Other issues were to keep a consistent temperature during the evaporation process of the sediment samples and keep away dust from entering the pot. Another method of removing the water from the sediment sample would be to let the sediment sample settle for at least 24 hours, and then remove the water by siphoning and centrifuging (Smith et al., 2018).

A limitation of the XRF analysis of the samples was that the samples were only performed once. The elements Ba, Sr, Mg, As, Cu, Sb, Cr, Ta, Co and Ni were removed from further analysis before the tracer screening process because they were not detected in a sample that was used. This issue might have been solved by more XRF measurements. Two soil samples were also excluded from the analysis because of a composition that deviated from the expected values.

About tracer screening process

The tracer screening process aims to select an optimal subset of tracers that maximize sample source discrimination while avoiding redundancy. The goal is to identify the smallest set of tracers that provides the most information about the differences between sources (Collins & Walling, 2002). However, other studies suggest that sediment fingerprinting research should strive to utilize the broadest possible range of tracers for source separation. This approach is limited only by the requirement to maintain conservative properties in those tracers, and employing a larger number of tracers, as opposed to a smaller set, leads to a reduction in errors when compared to synthetic mixtures (Smith, Karam, & Lennard, 2018).

The tracer screening process for this study resulted in selecting the elements Silicon (Si), Aluminum (Al), and Iron (Fe). This is a geochemically reasonable result, considering that these elements have a very high abundance in the upper continental crust by weight: Si (≈ 29.6%), Al (≈ 7.9%), and Fe (≈ 4.1%), with Oxygen (O) being the most abundant element at approximately 47.0% (Yaroshevsky, 2006).

Other elements with a relatively high abundance in the upper continental crust are Sodium (Na) (2.2%), Calcium (Ca) (2.9%), and Magnesium (Mg) (1.7%) (Yaroshevsky, 2006). These elements exhibit high solubility in water and are prone to dissolving when sediment is submerged, and they tend to react under varying environmental conditions (Chalaux-Clergue et al., 2024), which might be why they did not pass the screening process. Aluminum (Al) and Silicon (Si) show a lower tendency to react under varying environmental conditions, making them more reliable as tracers (Chalaux-Clergue et al., 2024).

5.4. Recommendations for Erosion Control Measures

To mitigate soil erosion in the Nyagako sub-catchment area, this study suggests implementing targeted soil conservation measures in the most erosion-prone area, Zone C, to reduce sediment contribution at the outlet. Furthermore, the promotion of reforestation and sustainable agricultural practices could help balance the cultivation-forest ratio, leading to the mitigation of soil erosion in cultivated areas. Additionally, regulating and monitoring mining activities in sub-zones A1 and A5 could lead to reduced turbidity and sediment contribution. Lastly, address the significance contribution of CA by stabilizing the channelized watercourses and slope bases.

5.5. Recommendations for Erosion Analysis and further Research

For future assessments on soil erosion, this study recommends enhancing GIS and turbidity monitoring, and to continue using GIS analysis for detailed mapping and monitoring of

erosion-prone areas. Furthermore, turbidity measurement programs should be expanded to track and mitigate the impacts of rainfall events and land disturbances on soil erosion.

When conducting this study, several interesting themes and research areas for further investigations were identified. Recommendations for further research are to conduct additional studies to understand the unexplained GG contributions from Zone A, investigate the contamination from landfills affecting the river system, address the significant contributions from CA by further research on how to stabilize alluvium and colluvium deposits from erosion. Furthermore, evaluate erosion management strategies, and conduct fingerprinting analysis during and after rainfall events, and also collect more soil samples in the catchment area to get more detailed data to map the geological signature and erosion dynamics.

6. Conclusion

This study investigated the sources and contributing factors of soil erosion in the Nyagako sub-catchment by employing GIS-analysis, conducting a sediment fingerprinting analysis and measuring turbidity. Zone C (1589 hectare), the most downstream region, was identified as the most erosion prone zone, corresponding to approximately 70-80% of the contribution at the outlet of the study site.

The most contributing source per hectare throughout was the source corresponding to CA; 0.11% per hectare in Zone A, 0.47% per hectare in Zone B and 0.29% per hectare in Zone C. CA encompasses both alluvium and colluvium deposits and often accumulates in channelized watercourse systems and slope base, respectively. Zone A was identified as the second highest contributing region, where the source of GG represented the major contributor corresponding to approximately 60.9% of the contribution at the outlet of Zone A and 11.3% of contribution at the final outlet. Further analysis of the correlation between the general trend of the contributing factor did not motivate the relatively large contribution of GG from Zone A. However, the sub zones ‘A1’ and ’A5’ indicated high turbidity contribution, suggesting impact from mining activity. Another reason could be impact from the landfill in the zone, contaminating the river.

The major contributing factor was identified as the cultivation-forest ratio, where there was a strong positive correlation between the contribution per hectare and cultivation-forest ratio for the respective zones and source groups. On the other hand, the steepness gradient showed a negative correlation to the contribution per hectare, which could be explained by less cultivating activity or higher degree of forest plantation occurring in the steeper regions.

The study's methodology provided valuable insights into soil erosion dynamics by combining the different methods, enhancing result interpretation. Sediment fingerprinting revealed source contribution and the variability in different regions, illustrating the complexity of soil erosion. GIS analysis was crucial for preparatory investigation, project design, result

interpretation, and validation. The correlation between MixSIAR, turbidity measurements, and the land degradation map (RUSLE) validated the results of each method. In addition, turbidity analysis complemented the fingerprinting results by mapping sub-catchments – identifying anthropogenic disturbances – and highlighting rainfall impacts.

In summary, this study illustrates the value of a multi-faceted approach in assessing soil erosion. By integrating GIS analysis, sediment fingerprinting, and turbidity measurements, it achieves a comprehensive understanding of the contributing factors and erosion-prone regions and sources of soil erosion in the Nyagako sub-catchment. The findings emphasize the significant roles of land use and anthropogenic activities, such as mining, in influencing soil erosion. The methodologies and insights gained from this study can inform future soil conservation efforts and help in the strategic planning of erosion mitigation measures. Additionally, the study highlights the need for continued research and data collection to refine the understanding of soil erosion processes and improve the accuracy of predictive models.

7. References

Bryan, R. B. (2000). Soil erodibility and processes of water erosion on hillslope. *Geomorphology, 32*(3-4), 385-415. https://doi.org/10.1016/S0169-555X(99)00105-1

Byers III, A. C. (1992). Soil loss and sediment transport during the storms and landslides of May 1988 in Ruhengeri Prefecture, Rwanda. *Natural Hazards, 5*, 279-292. Kluwer Academic Publishers.

Chalaux-Clergue, T., Bizeul, R., Batista, P. V. G., Martínez-Carreras, N., Laceby, J. P., & Evrard, O. (2024). Sensitivity of source sediment fingerprinting to tracer selection methods. *SOIL, 10*(109-138). https://doi.org/10.5194/soil-10-109-2024

Collins, A. L., & Walling, D. E. (2002). Selecting fingerprint properties for discriminating potential suspended sediment sources in river basins. *Water Resources Research, 45*(2), 210-225. https://doi.org/10.1029/2008WR007398

Clarke, R. T., & Minella, J. P. G. (2016). Evaluating sampling efficiency when estimating sediment source contributions to suspended sediment in rivers by fingerprinting. *Hydrological Processes, 30*(22), 3408-3419. https://doi.org/10.1002/hyp.10866

Eekhout, J. P. C., & de Vente, J. (2022). Global impact of climate change on soil erosion and potential for adaptation through soil conservation. *Earth-Science Reviews, 226*, 103921. https://doi.org/10.1016/j.earscirev.2022.103921

Evrard, O., Batista, P.V.G., Company, J., Dabrin, A., Foucher, A., Frankl, A., García-Comendador, J., Huguet, A., Lake, N., Lizaga, I., Martínez-Carreras, N., Navratil, O., Pignol, C., Sellier, V. (2022). Improving the design and implementation of sediment fingerprinting studies: summary and outcomes of the TRACING 2021 Scientific School. *Journal of Soils and Sediments, 22*(1648–1661). https://doi.org/10.1007/s11368-022-03203-1

FAO. (2019). Soil erosion: The greatest challenge for sustainable soil management. Food and Agriculture Organization of the United Nations. http://www.fao.org/3/ca7477en/ca7477en.pdf

Food and Agriculture Organization of the United Nations & Intergovernmental Technical Panel on Soils. (2015). Status of the World's Soil Resources (SWSR) – Main Report. FAO. http://www.fao.org/3/a-i5199e.pdf

GLOWS-FIU. (2016). Sediment Fingerprinting in the Nile Nyabarongo Upper Catchment, Rwanda. Global Water for Sustainability Program, Florida International University.

Government of Rwanda & International Union for Conservation of Nature (IUCN). (2022). The State of Soil Erosion Control in Rwanda. Rwanda Water Resources Board & Embassy of the Kingdom of the Netherlands in Rwanda.

Haldar, S. K., & Tišljar, J. (2014). Sedimentary rocks. In *Introduction to Mineralogy and Petrology* (pp. 117-144). Elsevier. https://doi.org/10.1016/C2011-0-07417-7

Hopkins, J. B., III, & Ferguson, J. M. (2012). Estimating the diets of animals using stable isotopes and a comprehensive Bayesian mixing model. *PLoS ONE, 7*(1), e28478. https://doi.org/10.1371/journal.pone.0028478

Igwe, P.U., Nwezi, C.C., Echendu, J.E., Chukwunyere, I.C., & Okonkwo, N.J. (2017). Adaptations to Soil Erosion: A Review. *International Journal of Advanced Engineering, Management and Science (IJAEMS), 3*(12). https://dx.doi.org/10.24001/ijaems.3.12.6

Jain, S.K., & Singh, V.P. (2003). Reservoir sedimentation. In S.K. Jain & V.P. Singh (Eds.), *Water Resources Systems Planning and Management* (Vol. 51, pp. 681-741). Elsevier.

James, R., Amasi, A. I., Wynants, M., Nobert, J., Mtei, K. M., & Njau, K. (2023). Tracing the dominant sources of sediment flowing towards Lake Victoria using geochemical tracers and a Bayesian mixing model. *Journal of Soils and Sediments, 23*(1568–1580). https://doi.org/10.1007/s11368-023-03440-y

Julien, P. Y. (2010). *Erosion and sedimentation* (2nd ed.). Cambridge University Press.

Karamage, F., Zhang, C., Kayiranga, A., Shao, H., Fang, X., Ndayisaba, F., Nahayo, L., Mupenzi, C., & Tian, G. (2016). USLE-based assessment of soil erosion by water in the Nyabarongo River catchment, Rwanda. *International Journal of Environmental Research and Public Health, 13*(835). https://doi.org/10.3390/ijerph13080835

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

Kříbek, B., Majer, V., Veselovský, F., & Nyambe, I. (2010). Discrimination of lithogenic and anthropogenic sources of metals and sulphur in soils of the central-northern part of the Zambian Copperbelt Mining District: A topsoil vs. subsurface soil concept. *Journal of Geochemical Exploration, 104*(1), 69-86. https://doi.org/10.1016/j.gexplo.2009.12.005

Majoro, F., Wali, U. G., Munyaneza, O., Naramabuye, F.-X., Bugenimana, E. D., & Mukamwambali, C. (2020). Farmers perceptions on soil erosion in Sebeya catchment, Rwanda. *EPH - International Journal of Agriculture and Environmental Research, 6*(10). https://www.researchgate.net/publication/345308989

Majoro, F., Wali, U. G., Munyaneza, O., & Naramabuye, F.-X. (2023). Sustainability Analysis of Soil Erosion Control in Rwanda: Case Study of the Sebeya Watershed. *Sustainability, 15*(3), 1969. https://doi.org/10.3390/su15031969

Marzen, M., & Iserloh, T. (2021). Processes of raindrop splash and effects on soil erosion. In J. Rodrigo-Comino (Ed.), *Precipitation: Earth Surface Responses and Processes* (pp. 351-371). University of Granada.

McGehee, R. P., Flanagan, D. C., Srivastava, P., & Nearing, M. A. (2021). Rainfall erosivity: Essential historical conceptual and practical perspectives for continued application. In *Precipitation* (Chapter 16). Elsevier. https://doi.org/10.1016/B978-0-12-822699-5.00014-8

Miller, B. A., & Juilleret, J. (2020). The Colluvium and Alluvium Problem: Historical Review and Current State of Definitions. *Earth-Science Reviews*. https://doi.org/10.1016/j.earscirev.2020.103362

Nahayo, A., Pan, G., & Joseph, S. (2016). Factors influencing the adoption of soil conservation techniques in Northern Rwanda. *Journal of Plant Nutrition and Soil Science, 179*(3), 367–375. https://doi.org/10.1002/jpln.201500403

Owens, P. N., Blake, W. H., Gaspar, L., Gateuille, D., Koiter, A. J., Lobb, D. A., Petticrew, E. L., Reiffarth, D. G., Smith, H. G., & Woodward, J. C. (2016). Fingerprinting and tracing the sources of soils and sediments: Earth and ocean science, geoarchaeological, forensic, and human health applications. *Earth-Science Reviews, 162*, 1–23. https://doi.org/10.1016/j.earscirev.2016.08.012

PKabatha. (2023). Rwanda Land Degradation Monitoring and Assessment 2022 [Map Image Layer]. Retrieved from https://www.arcgis.com/home/item.html?id=3d03f481508d4cf7965d2c6b6ab97d53

Prieto Tejedor, J. (Ed.). (2017). *Bayesian Inference*. IntechOpen. https://dx.doi.org/10.5772/66264

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Smith, H. G., Karam, D. S., & Lennard, A. T. (2018). Evaluating tracer selection for catchment sediment fingerprinting. *Journal of Soils and Sediments, 18*(3005–3019). https://doi.org/10.1007/s11368-018-1990-7

Stock, B. C., Jackson, A. L., Ward, E. J., Parnell, A. C., Phillips, D. L., & Semmens, B. X. (2018). Analyzing mixing systems using a new generation of Bayesian tracer mixing models. *PeerJ, 6*, e5096. https://doi.org/10.7717/peerj.5096

Svoray, T. (2022). *A Geoinformatics Approach to Water Erosion: Soil Loss and Beyond*. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-91536-0

Wilkinson, S. N., Hancock, G. J., Bartley, R., Hawdon, A. A., & Keen, R. J. (2013). Using sediment tracing to assess processes and spatial patterns of erosion in grazed rangelands, Burdekin River basin, Australia. *Agriculture, Ecosystems & Environment, 180*, 90–102. https://doi.org/10.1016/j.agee.2012.02.002

Yaroshevsky, A. A. (2006). Abundances of chemical elements in the Earth's crust. *Geochemistry International, 44*(1), 48-55. https://doi.org/10.1134/S001670290601006X

Zeghmar, A., Mokhtari, E., & Marouf, N. (2024). A machine learning approach for RUSLE-based soil erosion modeling in Beni Haroun dam Watershed, Northeast Algeria. *Earth Science Informatics*. https://doi.org/10.1007/s12145-024-01305-7

Appendix

A.1 XRF-Analysis Result

The provided tables present raw data obtained from the XRF analysis. Part A in Table A1.1a includes elements with the most significant results, which underwent the tracer screening process. Part B in table A1.1b lists elements that did not yield significant results, with "not detected" (ND) values recorded. Samples SS20 and SS2 were not used in the final analysis because they deviated from expected values..

Table A1.1a. PartA: Sorted raw dataset from the X-ray Fluorescence (XRF) measurements.

Sample id	Zone	Source	Si	Al	Fe	K	Ti	Ca	P	Zr	Rb	Nb	Mn	Zn	Pb
SS14	A	CA	22.025	13.265	10.332	3.035	1.029	0.191	0.101	0.089	0.031	0.009	0.032	0.011	0.008
ST17	A	CA	21.784	12.358	13.492	1.986	0.936	0.484	0.200	0.089	0.032	0.017	0.180	0.018	0.015
SS12	A	GG	21.488	15.731	8.544	2.857	0.580	0.219	0.105	0.046	0.049	0.009	0.050	0.010	0.009
SS15	A	GG	21.632	17.573	7.373	1.794	0.566	0.158	0.073	0.044	0.019	0.007	0.024	0.005	0.006
ST16	A	CA	22.928	12.137	9.816	2.746	1.496	0.315	0.172	0.126	0.032	0.014	0.093	0.011	0.009
SS11	A	GG	20.357	14.147	11.420	2.623	0.511	0.363	0.150	0.037	0.037	0.009	0.039	0.014	0.007
ST13	A	GG	21.445	14.241	9.827	2.691	0.781	0.244	0.112	0.066	0.043	0.009	0.110	0.012	0.008
SS18	A	CA	23.098	11.840	10.582	2.049	0.643	0.245	0.044	0.039	0.031	0.009	0.033	0.015	0.008
ST19	A	GG	21.106	14.160	9.105	2.642	2.027	0.313	0.126	0.033	0.015	0.007	0.081	0.013	0.007
*SS20**	A	GG	24.511	18.775	2.083	2.143	0.000	0.188	0.053	0.008	0.038	0.005	0.009	0.005	0.006
SS21	A	GG	19.857	16.936	10.718	1.993	0.463	0.221	0.120	0.036	0.025	0.006	0.040	0.008	0.011
SS22	A	QS	19.309	15.562	12.692	2.075	1.133	0.198	0.086	0.073	0.020	0.008	0.044	0.012	0.007
SS23	A	QS	19.180	15.515	13.601	1.954	1.105	0.216	0.103	0.075	0.019	0.010	0.043	0.009	0.007
Up6_C1	Node1	–	20.399	16.929	8.079	2.726	0.288	0.745	0.135	0.014	0.051	0.005	0.095	0.012	0.011
Up6_C2	Node1	–	20.564	16.751	8.326	2.514	0.076	0.634	0.250	0.015	0.041	0.004	0.058	0.013	0.013
Up6_C3	Node1	–	21.799	18.692	5.679	1.724	0.161	0.442	0.099	0.013	0.039	0.003	0.046	0.006	0.007
ST5	B	GG	22.044	15.373	9.168	1.655	0.748	0.340	0.133	0.088	0.037	0.012	0.075	0.011	0.011
SS6	B	GG	20.596	17.875	9.796	0.783	0.373	0.136	0.084	0.042	0.025	0.010	0.036	0.006	0.010
SS7	B	GG	21.446	17.728	8.804	0.880	0.393	0.147	0.079	0.039	0.023	0.009	0.035	0.006	0.007
ST8	B	GG	21.471	15.928	8.573	1.707	0.903	0.267	0.128	0.049	0.030	0.012	0.102	0.012	0.009
SS9	B	GG	22.697	17.123	5.420	2.455	0.472	0.179	0.052	0.037	0.038	0.005	0.036	0.006	0.004
SS10	B	GG	21.634	14.301	10.947	1.614	1.139	0.231	0.147	0.072	0.022	0.013	0.069	0.011	0.009
SS27	B	QS	19.801	14.682	12.877	2.925	0.659	0.201	0.242	0.059	0.040	0.006	0.030	0.006	0.012
ST29	B	GG	22.581	13.726	8.365	2.898	0.599	0.600	0.165	0.074	0.044	0.010	0.099	0.015	0.009
Stream6_C1	Node2	–	18.998	11.900	11.583	3.990	0.407	1.434	0.246	0.024	0.060	0.009	0.217	0.019	0.009
Stream6_C2	Node2	–	20.205	13.240	10.291	2.909	0.449	1.817	0.112	0.024	0.044	0.008	0.172	0.023	0.010
Stream6_C3	Node2	–	20.286	11.051	9.547	3.231	0.306	2.238	0.246	0.020	0.044	0.008	0.223	0.022	0.008
Mix6_C1	Node3	–	20.514	14.169	8.553	2.690	0.345	1.471	0.170	0.019	0.045	0.006	0.150	0.015	0.010
Mix6_C2	Node3	–	20.706	16.699	8.553	2.127	0.197	0.728	0.216	0.017	0.039	0.004	0.076	0.010	0.009
Mix6_C3	Node3	–	22.239	17.530	5.461	1.987	0.127	0.624	0.108	0.014	0.037	0.004	0.053	0.008	0.007
ST1	C	QS	18.002	13.632	16.408	3.939	0.810	0.264	0.150	0.051	0.043	0.008	0.019	0.012	0.008
*SS2**	C	QS	24.588	15.081	3.367	5.024	0.494	0.174	0.086	0.072	0.036	0.003	ND	ND	0.004
SS3	C	QS	17.225	15.747	14.794	4.106	0.358	0.192	0.155	0.033	0.041	0.004	0.009	0.010	0.010
SS4	C	QS	18.407	15.325	15.583	2.368	0.376	0.149	0.121	0.032	0.023	0.003	0.022	0.006	0.003
SS24	C	CA	21.426	13.216	11.393	4.561	0.470	0.220	0.134	0.046	0.066	0.006	0.045	0.016	0.003
SS30	C	CA	23.842	11.889	8.036	2.901	1.435	0.374	0.138	0.112	0.034	0.011	0.332	0.015	0.008
SS31	C	CA	24.706	12.933	7.048	2.836	1.007	0.224	0.102	0.103	0.028	0.008	0.088	0.007	0.007
SS32	C	QS	20.447	13.621	12.994	2.369	1.134	0.200	0.109	0.079	0.026	0.010	0.087	0.010	0.009
ST33	C	CA	22.247	10.493	11.272	2.940	1.224	0.443	0.217	0.163	0.026	0.012	0.109	0.015	0.009
ST25	C	QS	18.460	12.366	12.120	2.670	0.783	0.543	0.347	0.045	0.051	0.015	0.079	0.022	0.011
SS26	C	GG	22.316	18.816	4.642	2.091	0.055	0.193	0.056	0.017	0.037	0.005	0.016	0.006	0.004
SS28	C	QS	20.076	15.462	12.171	2.500	0.664	0.155	0.222	0.056	0.040	0.005	0.029	0.005	0.007
Mix11_C1	Node4	–	20.234	14.076	11.503	3.014	0.875	0.400	0.181	0.044	0.043	0.010	0.117	0.013	0.008
Mix11_C2	Node4	–	20.763	13.592	10.020	2.827	0.258	1.469	0.171	0.022	0.041	0.007	0.165	0.023	0.012
Mix11_C3	Node4	–	21.495	15.939	7.406	2.238	0.194	0.973	0.111	0.019	0.032	0.005	0.109	0.011	0.006
Mix11_SC	Node4	–	20.319	12.920	10.529	3.311	0.458	1.687	0.154	0.027	0.046	0.007	0.188	0.017	0.007

*Samples SS20 and SS2 was not used in the final analysis

Table A1.1b. Part B: Sorted raw dataset from the X-ray Fluorescence (XRF) measurements.

Sample id	Zone	Source	Ba	Sr	Mg	As	Cu	Sb	Cr	Ta	Co	Ni
SS14	A	CA	0.065	0.010	0.903	0.003	0.004	0.008	0.053	0.006	0.000	0.007
ST17	A	CA	0.060	0.011	0.000	ND	0.005	0.015	0.052	ND	0.016	ND
SS12	A	GG	0.057	0.006	0.828	0.000	ND	0.010	0.039	0.009	0.000	ND
SS15	A	GG	0.054	0.006	0.493	ND	ND	0.009	ND	0.006	0.010	ND
ST16	A	CA	0.055	0.009	0.882	0.003	0.004	0.010	0.033	0.008	0.000	0.006
SS11	A	GG	0.083	0.005	1.637	0.000	0.007	ND	0.051	ND	0.000	0.009
ST13	A	GG	0.067	0.008	1.417	0.001	ND	0.006	ND	ND	ND	ND
SS18	A	CA	0.057	0.009	0.415	0.008	ND	0.010	0.041	0.010	0.000	ND
ST19	A	GG	0.056	0.006	1.131	0.003	0.000	0.010	0.058	0.012	0.000	0.010
SS20	A	GG	ND	0.002	0.000	0.000	ND	ND	ND	0.004	0.000	ND
SS21	A	GG	0.053	0.008	0.631	0.003	0.006	ND	0.046	0.007	0.006	0.007
SS22	A	QS	0.048	0.004	0.699	0.002	0.008	0.006	0.048	0.009	0.000	0.011
SS23	A	QS	0.056	0.004	ND	0.003	0.004	0.006	0.048	0.008	0.007	0.006
Up6_C1	Node1	–	0.051	0.026	0.850	0.004	0.008	0.006	0.044	ND	0.000	0.008
Up6_C2	Node1	–	0.056	0.041	1.180	0.004	0.011	0.006	ND	0.011	0.006	0.013
Up6_C3	Node1	–	0.031	0.014	0.368	0.004	0.006	ND	ND	0.009	0.004	0.006
ST5	B	GG	0.033	0.008	0.578	0.000	ND	0.012	0.031	0.007	0.000	ND
SS6	B	GG	0.036	0.005	0.462	0.000	ND	0.006	ND	0.007	0.006	0.005
SS7	B	GG	0.030	0.003	ND	ND	0.004	0.004	ND	0.008	0.000	ND
ST8	B	GG	0.050	0.005	1.023	0.000	ND	ND	0.033	0.006	ND	0.007
SS9	B	GG	0.049	0.004	0.797	0.001	ND	0.009	0.031	0.007	0.000	0.000
SS10	B	GG	0.049	0.006	0.633	0.002	ND	0.012	0.040	0.010	0.000	ND
SS27	B	QS	0.080	0.032	0.509	0.005	0.011	ND	0.044	0.000	0.007	ND
ST29	B	GG	0.053	0.009	1.266	0.000	0.000	0.006	ND	ND	0.000	0.008
Stream6_C1	Node2	–	0.059	0.023	2.535	0.008	0.009	0.009	0.072	0.012	0.015	0.016
Stream6_C2	Node2	–	0.073	0.024	0.995	0.007	0.022	0.016	ND	ND	0.000	0.014
Stream6_C3	Node2	–	0.071	0.032	4.137	0.003	0.013	0.012	0.044	0.014	0.014	0.007
Mix6_C1	Node3	–	0.055	0.028	2.341	0.004	0.009	0.006	0.022	0.010	0.007	0.007
Mix6_C2	Node3	–	0.049	0.044	0.779	0.007	0.009	0.006	ND	0.008	0.000	0.010
Mix6_C3	Node3	–	0.033	0.015	0.798	0.004	0.006	0.008	ND	0.009	0.000	0.005
ST1	C	QS	0.077	0.014	ND	0.025	ND	0.010	0.059	0.011	0.000	0.013
SS2	C	QS	0.067	0.009	0.339	0.004	ND	0.007	0.052	0.006	ND	0.006
SS3	C	QS	0.063	0.030	0.756	0.034	ND	0.007	0.042	ND	0.000	0.007
SS4	C	QS	0.052	0.011	0.520	0.001	ND	0.009	0.053	0.007	0.000	0.009
SS24	C	CA	0.079	0.004	0.444	0.006	0.006	0.005	0.035	0.011	ND	0.006
SS30	C	CA	0.054	0.011	1.262	0.002	0.008	0.011	0.037	ND	0.010	ND
SS31	C	CA	0.038	0.007	0.675	0.002	0.005	0.005	ND	0.006	0.009	ND
SS32	C	QS	0.068	0.006	0.877	0.004	0.007	0.009	0.043	0.007	0.005	0.011
ST33	C	CA	0.071	0.008	2.399	0.003	0.009	0.011	0.056	ND	ND	ND
ST25	C	QS	0.074	0.006	4.471	0.000	0.000	0.024	0.081	0.018	0.022	0.013
SS26	C	GG	ND	ND	0.704	ND	0.000	0.007	0.032	0.004	0.005	ND
SS28	C	QS	0.074	0.029	ND	0.007	0.010	0.007	0.033	0.006	0.000	ND
Mix11_C1	Node4	–	0.064	0.010	1.104	0.001	0.007	0.009	0.050	0.011	0.000	0.007
Mix11_C2	Node4	–	0.075	0.032	1.415	0.005	0.015	0.008	0.062	0.012	0.000	0.012
Mix11_C3	Node4	–	0.040	0.022	1.152	0.007	0.011	0.008	ND	0.012	0.000	ND
Mix11_SC	Node4	–	0.048	0.025	1.386	0.007	0.012	0.009	0.055	0.009	0.000	0.006

A.1.1 Tracer Screening Process Result

The raw geochemical dataset of the samples was sorted into potential source groups by location and different lithology as presented in table A1.1a-b . The elements were sorted from the ones that showed high concentration performance and the elements that were not detected were removed. The elements that were removed were; Barium (Ba), Strontium (Sr), Magnesium (Mg), Arsenic (As), Copper (Cu), Antimony (Sb), Chromium (Cr), Tantalum (Ta), Cobalt (Co), and Nickel (Ni). The other remaining elements were Silicon (Si), Aluminum (Al), Iron (Fe), Potassium (K), Titanium (Ti), Calcium (Ca), Phosphorus (P), Zirconium (Zr), Rubidium (Rb), Niobium (Nb), Zinc (Zn), and Lead (Pb). The remaining elements were further analyzed through the 'range test', see figure A1.1.

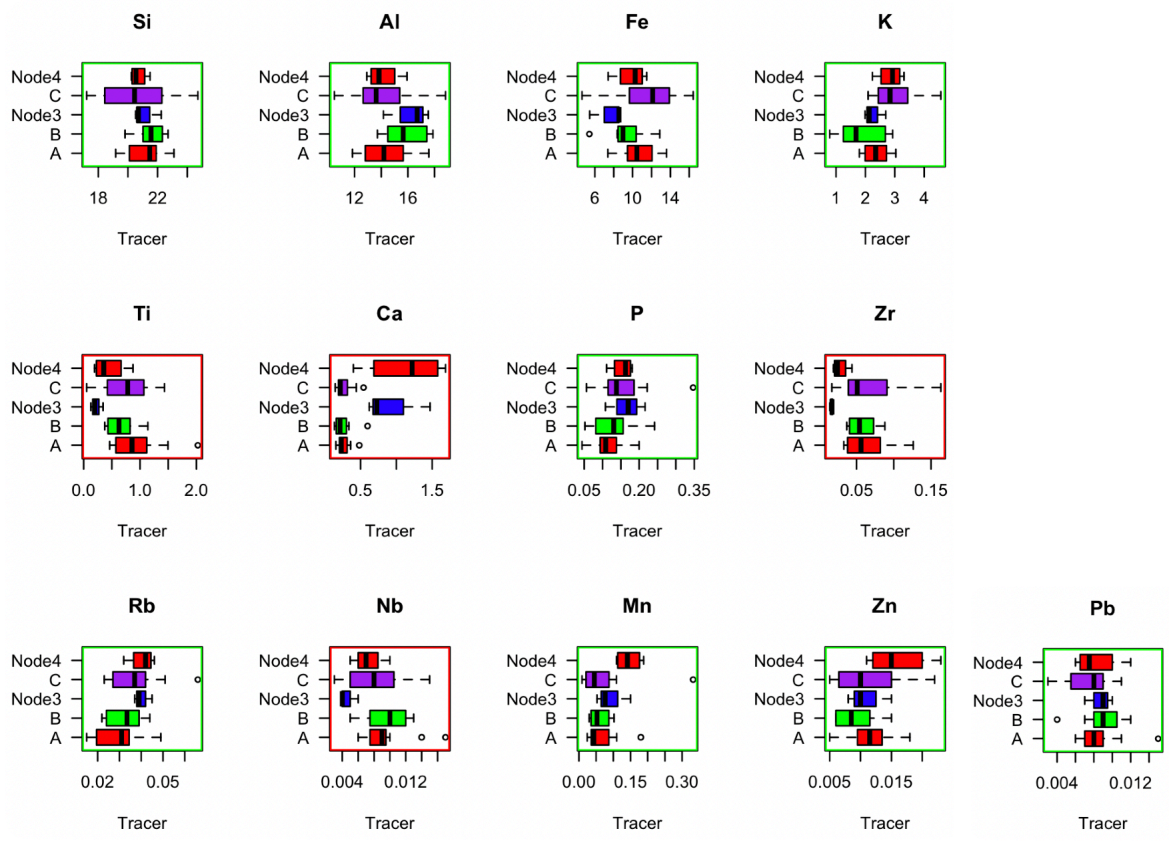

Figure A1.1. Range test result: If the tracer element is conservative then the frame is green, it is when the downstream node value is in the concentration range of the upstream samples, otherwise it is red and thereby non-conservative. A represents samples from Zone A, B from Zone B and C from Zone C.

The result from the range test failed the elements; Ti, Ca, Zr and Nb. The analysis of these elements indicated a non-conservative behavior and were thereby not analyzed further.

The Kruskal-Wallis test results for the remaining elements are presented in table A1.2. For all elements, where the p-values are less than 0.05 (highlighted with green), are suggesting that there are statistically significant differences in median values among the sources, indicate significant differences in the distributions of Si, Al, Fe, K and Mn across the sources CA, GG and QS, especially for Si, Al and Fe. In contrast, P, Rb, Zn and Pb are not showing any distinct distribution between the sources (p-value greater than 0.05, highlighted in red).

Table A1.2. Results for evaluation of differences in distribution between the sources, based on the Kruskal-Wallis test.

Element	ChiSquared	DF	P-Value
Si	20.55	2	0.000035
Al	18.12	2	0.000117
Fe	18.12	2	0.000116
K	7.171	2	0.027726
P	3.811	2	0.148753
Rb	0.290	2	0.865050
Mn	6.039	2	0.048820
Zn	5.706	2	0.057661
Pb	0.041	2	0.979946

The intrasource variance of the remaining elements, as detailed in Table A1.3, illustrates the variability for each element in different sources (CA, GG, and QS). Higher values indicate a greater degree of variability for the given element associated with a particular source, pointing to a broader range of concentrations. Notably, Iron (Fe) displays significant variability within sources, particularly in CA and GG, which suggests diverse concentrations within these source categories.

Table A1.3. The intrasource variance of each tracer within each source.

Source	Si	Al	Fe	K	Mn
CA	1.23	0.83	4.05	0.62	0.0102
GG	0.66	2.88	3.73	0.47	0.0010
QS	1.10	1.40	2.39	0.59	0.0007

The intersource variance of the remaining elements in table A1.4 highlights the average variability of each element across the different sources (CA, GG, and QS). High values, like those for Fe, suggest this element's concentration significantly varies between sources, which could be useful for distinguishing between them. Low values indicate less variability between sources, potentially making it less useful for differentiating the sources.

Table A1.4. The intersource variance summary.

Element	Intersource Variance
Si	2.45
Al	2.47
Fe	4.37
K	0.145
Mn	0.001

Due to their higher intrasource variance relative to the intersource variance, Potassium (K) and Manganese (Mn) were excluded from further analysis. The elements that continue to be analyzed as tracers are Silicon (Si), Aluminum (Al), and Iron (Fe).

The result of the Shapiro-Wilk test for Si, Al, and Fe shows no significant evidence to reject the null hypothesis of normality for any of the tracers (table A1.5). The p-value is well above the common alpha level of 0.05, indicating no significant deviation from normality. The distribution of tracers values is considered to be normally distributed. This is conducive to analyses or models assuming normality of data (Stock et al. 2018).

Table A1.5. The result of the Shapiro-Wilk test.

Tracer	W-Statistic	p-value
Si	0.91298	0.2014
Al	0.95550	0.6835
Fe	0.94534	0.5296

The W statistic in the Shapiro-Wilk test shows the measure of normality, assessing how well a dataset conforms to a normal distribution. All of the tracer's seem to resemble a normal distribution.

A. 1.2 Sampling Locations

The following tables present the sampling locations for the sediment samples in table A1.6, soil samples in table A1.7, and TSS and turbidity measurements in table A1.8.

Table A1.6. Sediment sample collections, time and location.

Sampling point	Mixture	Location		Dates of Collection			
		Easting	**Northing**	***2023-10-18***	***2023-11-11***	***2023-11-29***	***2023-12-10***
Upstream 6	Node 1	467373	4773351	X	X	X	–
Stream 6	Node 2	467396	4773342	X	X	X	–
Mix 6	Node 3	467381	4773429	X	X	X	–
Mix 11	Node 4	463682	4774407	–	–	–	X

Table A1.7. Soil sample collection, horizon, location and date.

			Location		
Sample id	**Collection id**	**Soil Horizon**	**Easting**	**Northing**	**Collected**
ST1	*S_top1*	topsoil	468256	4776222	2023-11-29
SS2	*S_sub2*	subsoil	468076	4776763	2023-11-29
SS3	*S_sub3*	subsoil	468232	4776241	2023-11-29
SS4	*S_sub4*	subsoil	468412	4776236	2023-11-29
ST5	*S_top5*	topsoil	469930	4774895	2023-11-29
SS6	*S_sub6*	subsoil	469916	4774826	2023-11-29
SS7	*S_sub7*	subsoil	470290	4774036	2023-11-29
ST8	*S_top8*	topsoil	470348	4774026	2023-11-29
SS9	*S_sub9*	subsoil	470465	4773994	2023-11-29
SS10	*S_sub10*	subsoil	469573	4772412	2023-11-29
SS11	*S_sub11*	subsoil	470037	4771941	2023-11-29
SS12	*S_sub12*	subsoil	469519	4771202	2023-11-29
ST13	*S_top13*	topsoil	469258	4769898	2023-11-29
SS14	*S_sub14*	subsoil	469319	4769630	2023-12-10
SS15	*S_sub15*	subsoil	469293	4769904	2023-12-10
ST16	*S_top16*	topsoil	469374	4769666	2023-10-19
ST17	*S_top17*	topsoil	469294	4769638	2023-10-19
SS18	*S_sub18*	subsoil	469253	4769879	2023-11-29
ST19	*S_top19*	topsoil	468525	4770312	2023-10-19
SS20	*S_sub20*	subsoil	466765	4771340	2023-12-10
SS21	*S_sub21*	subsoil	466239	4771504	2023-11-29
SS22	*S_sub22*	subsoil	467040	4773270	2023-12-10
SS23	*S_sub23*	subsoil	467048	4773267	2023-10-19
SS24	*S_sub24*	subsoil	467362	4773466	2023-12-10
ST25	*S_top25*	topsoil	464996	4772324	2023-11-29
SS26	*S_sub26*	subsoil	463738	4773604	2023-11-29
SS27	*S_sub27*	subsoil	469668	4775759	2023-11-29
SS28	*S_sub28*	subsoil	463841	4774989	2023-11-29
ST29	*S_top29*	topsoil	469573	4772412	2023-11-29
SS30	*S_sub30*	subsoil	466527	4774354	2023-12-10
SS31	*S_sub31*	subsoil	465851	4774586	2023-12-10
SS32	*S_sub32*	subsoil	464977	4774603	2023-12-10
ST33	*S_top33*	topsoil	466471	4774432	2023-10-19

Table A1.8. Turbidity measurements, time and location.

Date:	Campaign 1		Campaign 2		Campaign 3		Location	
	2023-10-18		*2023-11-11*		*2023-12-10*			
Sampling point	**Turbidity (NTU)**	**TSS (mg/l)**	**Turbidity (NTU)**	**TSS (mg/l)**	**Turbidity (NTU)**	**TSS (mg/l)**	**Easting**	**Northing**
Upstream 1	3120	2920	–	–	3537	3759	469364	4769646
Stream 1	26	13	–	–	54	42	469328	4769640
Mix 1	1340	1630	–	–	2202	1984	469341	4769662
Stream 2	58	37	–	–	24	15	469283	4769872
Mix 2	–	–	–	–	854	905	469239	4769890
Stream 3	35	28	–	–	15	10	468539	4770252
Mix 3	–	–	–	–	1565	1622	468532	4770291
Upstream 4	952.5	–	–	–	–	–	468328	4770710
Stream 4	1250	1090	–	–	1226	1232	468287	4770712
Mix 4	2530	2600	–	–	1999	1876	468280	4770753
Upstream 5	1450	–	–	–	–	–	468087	4770964
Stream 5	60	–	–	–	19	8	468106	4770985
Mix 5	900	910	–	–	1403	1490	468075	4770993
Upstream 6	950	877	1155	1095	1423	1510	467373	4773351
Stream 6	363	274	383	335	215	188	467396	4773342
Mix 6	763.5	703	1081	972	819	827	467381	4773429
Upstream 7	587	519	–	–	787	799	466510	4774121
Stream 7	289	253	–	–	40	36	466487	4774133
Mix 7	565	559	–	–	757	786	466491	4774172
Stream 8	92	84	–	–	33	28	465990	4774478
Mix 8	247	271	–	–	450	435	465886	4774523
Upstream 9	–	–	349	335	–	–	465930	4774499
Stream 9	445	419	1833	1184	37	28	464843	4774639
Mix 9	–	–	–	–	501	477	464656	4774514
Upstream 10	10280	9080	–	–	–	–	464224	4774416
Stream 10	7710	7700	74	64	44	35	464212	4774400
Mix 10	7320	6030	262	252	412	431	464183	4774427
Mix 11	9840	8400	409	380	526	512	463682	4774407

A.2 MixSIAR Model Input Data

The following tables present the input data for the MixSIAR modeling, according to the sample design and MixSIAR framework. Source inputs are presented in table A2.1, mixture inputs in table A2.2 and discriminant input in table A2.3. The mean value and standard deviation of the samples representing upstream sources of the corresponding node. Note that *CA** and *QS**, in model 2 and *GG** in model 4 are using a generalized mean value, based on all of the samples in the catchment that represent that specific source.

Table A2.1. Source input for the MixSIAR models.

Source.m1.Geo5.csv	**Source**	**MeanSi**	**SDSi**	**MeanAl**	**SDAl**	**MeanFe**	**SDFe**	***n***
Model 1	*CA*	22.459	0.651	12.400	0.614	11.055	1.655	*4*
	GG	20.981	0.716	15.464	1.524	9.498	1.475	*6*
	QS	19.245	0.092	15.539	0.033	13.146	0.643	*2*
Source.m2.Geo5.csv	**Source**	**MeanSi**	**SDSi**	**MeanAl**	**SDAl**	**MeanFe**	**SDFe**	***n***
Model 2	*CA**	22.757	1.111	12.666	0.913	10.246	2.011	*8*
	GG	21.781	0.780	16.008	1.109	8.725	1.703	*5*
	*QS**	18.990	1.027	14.657	1.329	13.693	1.735	*9*
Source.m3.Geo5.csv	**Source**	**MeanSi**	**SDSi**	**MeanAl**	**SDAl**	**MeanFe**	**SDFe**	***n***
Model 3	*mix.1*	20.921	0.765	17.457	1.073	7.361	1.462	*3*
	mix.2	19.830	0.721	12.064	1.104	10.473	1.030	*3*
Source.m4.Geo5.csv	**Source**	**MeanSi**	**SDSi**	**MeanAl**	**SDAl**	**MeanFe**	**SDFe**	***n***
Model 4	*mix.3*	21.153	0.945	16.133	1.750	7.522	1.785	*3*
	CA	23.055	1.489	12.133	1.233	9.438	2.226	*4*
	*GG**	21.477	0.686	15.975	1.186	8.764	1.966	*14*
	QS	18.769	1.243	14.359	1.351	14.012	1.835	*6*

***** Synthetic source, meaning the average of all of the samples corresponding to the particular source was used, due to lack of samples representing the source in the particular zone if the number of samples was ($n \leq 1$).

Table A2.2. The mixture input for the MixSIAR models.

mix.1.Comb.copy.csv	**Sample ID**	**Campaign**	**Si**	**Al**	**Fe**
Model 1	Up6_C1	1	20.399	16.929	8.079
	Up6_C2	2	20.564	16.751	8.326
	Up6_C3	3	21.799	18.692	5.679
mix.2.Comb.copy.csv	**Sample ID**	**Campaign**	**Si**	**Al**	**Fe**
Model 2	Stream6_C1	1	18.998	11.900	11.583
	Stream6_C2	2	20.205	13.240	10.291
	Stream6_C3	3	20.286	11.051	9.547
mix.3.Comb.copy.csv	**Sample ID**	**Campaign**	**Si**	**Al**	**Fe**
Model 3	Mix6_C1	1	20.514	14.169	8.553
	Mix6_C2	2	20.706	16.699	8.553
	Mix6_C3	3	22.239	17.530	5.461
mix.4.Comb.copy.csv	**Sample ID**	**Campaign**	**Si**	**Al**	**Fe**
Model 4	Mix11_C1*	1	20.234	14.076	11.503
	Mix11_C2	2	20.763	13.592	10.020
	Mix11_C3	3	21.495	15.939	7.406
	Mix11_SC	SC**	20.319	12.920	10.529

* Collected after a rainfall event, which might affect the result.

** SC refers to the 'soil campaign', which could affect the result.

Table A2.3. Discriminant input for the MixSIAR models.

discr.m1.Geo5.csv	**Source**	**MeanSi**	**SDSi**	**MeanAl**	**SDAl**	**MeanFe**	**SDFe**	***n***
Model 1	*CA*	0	0	0	0	0	0	*4*
	GG	0	0	0	0	0	0	*6*
	QS	0	0	0	0	0	0	*2*
discr.m2.Geo5.csv	**Source**	**MeanSi**	**SDSi**	**MeanAl**	**SDAl**	**MeanFe**	**SDFe**	***n***
Model 2	*CA*	0	0	0	0	0	0	*8*
	GG	0	0	0	0	0	0	*5*
	QS	0	0	0	0	0	0	*9*
discr.m3.Geo5.csv	**Source**	**MeanSi**	**SDSi**	**MeanAl**	**SDAl**	**MeanFe**	**SDFe**	***n***
Model 3	*mix.1*	0	0	0	0	0	0	*3*
	mix.2	0	0	0	0	0	0	*3*
discr.m4.Geo5.csv	**Source**	**MeanSi**	**SDSi**	**MeanAl**	**SDAl**	**MeanFe**	**SDFe**	***n***
Model 4	*mix.3*	0	0	0	0	0	0	*3*
	CA	0	0	0	0	0	0	*4*
	GG	0	0	0	0	0	0	*14*
	QS	0	0	0	0	0	0	*6*