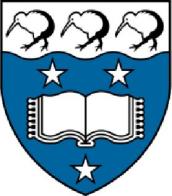
+

River Visualization



Yifan Wang

School of Computer Science / Department of Statistics

The University of Auckland

Supervisor: Charlotte Jones-Todd

A dissertation submitted in partial fulﬁllment of the requirements for the degree of Master of Professional Studies in Data Science / Master of Data Science, The University of Auckland, 20XX.

­­­­­­­

# Abstract

Put your abstract here. The abstract should contain a brief summary of the aim, methodologies, ﬁnding and conclusions of the dissertation. The abstract should normally be fewer than 350 words.

Table of Contents

Abstract 2

Chapter 1 Introduction 6

Chapter 2 Methodologies and analysis 7

Chapter 3 Discussion 8

Chapter 4 Conclusions 9

References 10

Appendix A Some extra things 11

# 1 Introduction

The River Environment Classification (REC) serves as a comprehensive database encapsulating the spatial attributes of catchment areas for every segment within New Zealand's extensive river network. The REC database encompasses over 30 attributes per river segment and typically includes spatial geographic data in formats such as shapefiles.

Traditionally, these types of data are processed and viewed using professional geographic information system (GIS) tools like QGIS. However, the use of such professional GIS tools requires extensive geographical knowledge and presents a steep learning curve.

To address this challenge, a visualization tool has been developed using the R programming language, specifically employing the Shiny framework to create an interactive web application. This approach allows for dynamic exploration and visually interpretation of the complex data contained within the REC and other water quality database. Through the Shiny application, specific river segments and attributes can be selected for visualization, facilitating insights into the connections and flows between river segments, as well as the spatial distribution and relationships of the catchment areas.

## Background

New Zealand's landscape is sculpted by an extensive network of rivers, which play a critical role in shaping the country's natural beauty and biodiversity. These rivers vary greatly in length, number of tributaries, and the topography they traverse, making them fascinating subjects of geographical and environmental studies.

The longest river in New Zealand is the Waikato River, which stretches approximately 425 kilometers from its source in the central North Island to its outlet into the Tasman Sea. This river, like many others in New Zealand, features a diverse array of tributaries, ranging from small streams to sizable secondary rivers that contribute to its flow and ecological diversity. For instance, the Clutha River in the South Island, the second longest in the country, is renowned for its extensive and complex system of tributaries that drain the Southern Alps.

The topography through which these rivers flow varies dramatically. The Waikato River begins its journey in the volcanic central plateau, meandering through lush farmland before cutting a path through the Hamilton lowlands. Conversely, rivers like the Shotover River in Otago, known for its dramatic and rugged scenery, carve through steep gorges and rocky terrains, influenced heavily by glacial and fluvial processes.

Understanding river networks requires familiarity with certain network and graph terminology, often used to describe the complex relationships and functions within these natural systems.

|  |  |
| --- | --- |
| Term | Definition |
| Source | The original point from which the river flows. |
| Edge | In river networks, edges are the stretches of river between two points (nodes), such as between two tributaries or between a tributary and the main river. | |
| Downstream Nodes | Points along the river closer to its mouth. They represent intersections or branching points within the river system moving towards the outlet. |
| Upstream Nodes | Points along the river closer to the source. They are the starting points of river flow within the network. |
| Path | A sequence of connected nodes and edges that define a route from one part of the river to another, typically from upstream to downstream. |
| Nearest Neighbours | Nodes directly connected to a given node by an edge without any other nodes in between. In the following thesis, all nearest neighbour are calculated by euclidean distance. |
| Connectivity | Describes how many nearest upstream nodes neighbours are connected to a certain stream |
| Outlet | The point where the river discharges its waters, typically into a sea, lake, or another river. Outlets are crucial for the ecological health of the surrounding areas, including estuaries and wetlands. |

New Zealand's rivers are vital ecological assets, characterized by their diverse lengths, tributaries, and the complex topographies they navigate. All these above making it an interesting object to visualization the rivers in New Zealand and reveal the topology structure and connectivity within these river networks.

## Motivation

There are existing methods for visualizing New Zealand's river systems, one method is offered by the National Institute of Water and Atmospheric Research (NIWA). NIWA’s NZ River Maps provides a overall information of all the river in New Zealand with a range of information select to show on the maps such as hydrology, bed sediment cover and etc [Figure 1]. Another method is Optimal Channel Network[1](OCN) proposed by Luca Carraro & Florian Altermatt, which can accurately show the appearance of the river in three-dimensional space and perform simple meta-population simulation within river network [Figure 2].

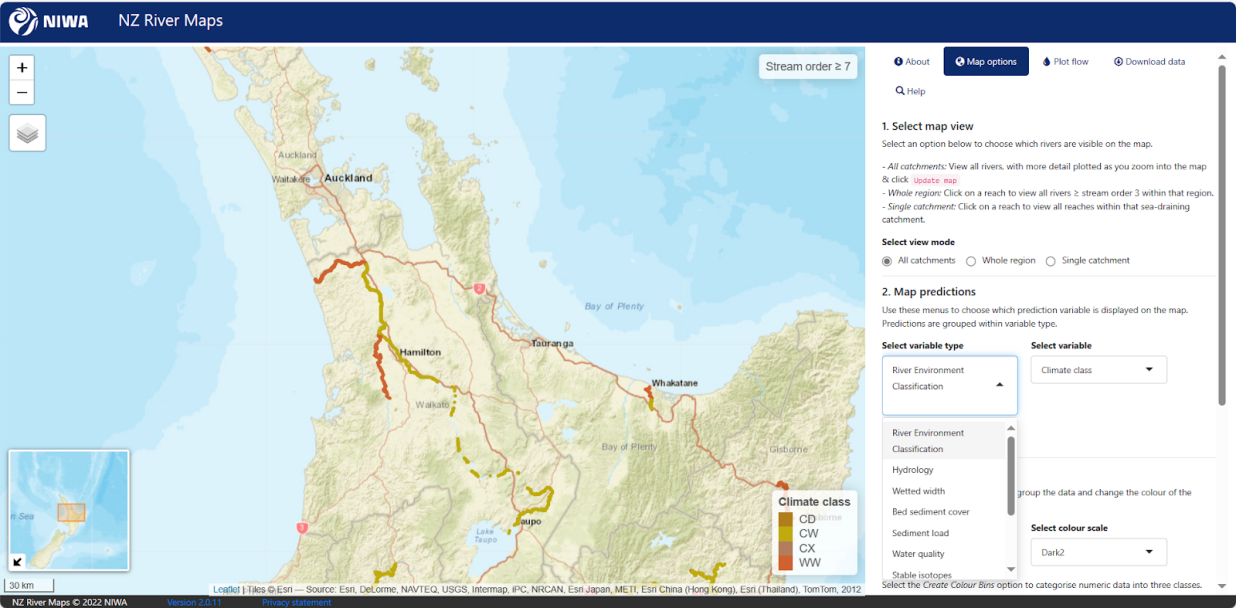


Figure 1. NZ River Maps of NIWA

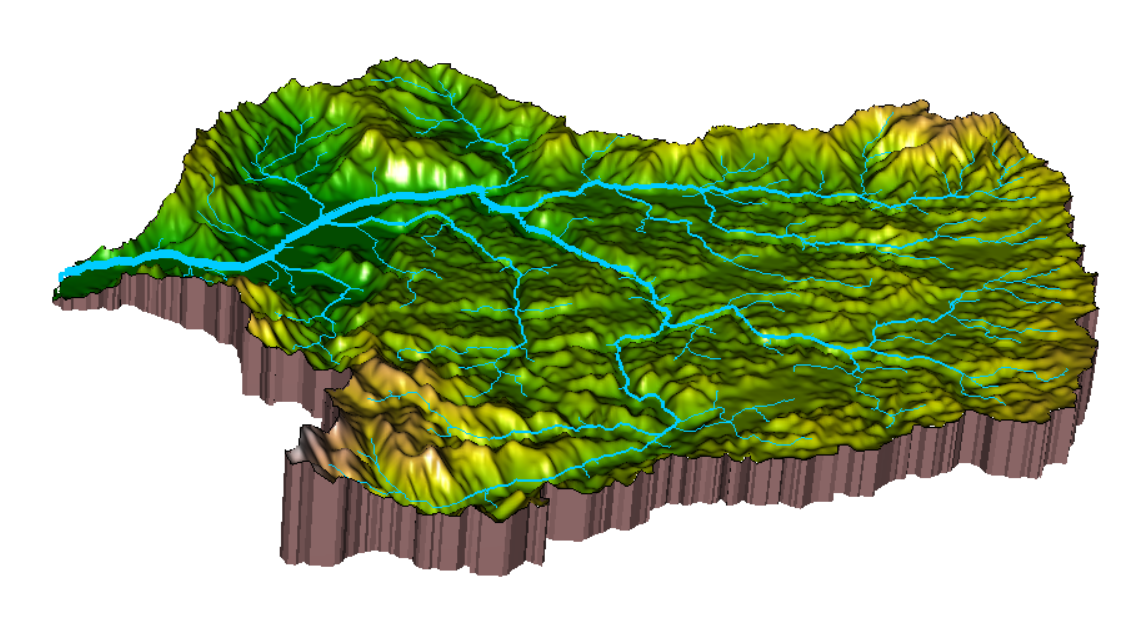


Figure 2. Visualisation of Optimal Channel Network Visualisation on Marokopa River

While approaches like above often fall short in effectively integrating and regionalizing water quality data to display pollution values comprehensively. Furthermore, they typically do not reveal the underlying topological logic of the river networks, which is crucial for understanding ecological connections and the flow dynamics within these systems. Meanwhile, it requires a number of geo-tools to get DEM data and knowledge of the WGS84 projection system coordinates of the river outlet in order to use OCN to generate realistic river images. In terms of the system's ease of operation, it is crucial to find a method that allows users to easily implement visualizations.

Recent years, the hierarchical characteristics of river network has been mentioned a lot as a critical attribute to reveal the topology of river network. The topology structure of the river network directly affects the transport processes of runoff, sediment and solutes [3], and is also a major control factor for many natural processes such as river greenhouse gas emissions, biodiversity, riparian vegetation function and food web structure [2]. The branch structure of river network is an important topological feature. The geometric and topological characteristics can quantitatively describe the relationship between hydrology, geomorphology, ecology and biological processes in river network, which is helpful for comprehensive numerical simulation and development of dynamic topological prediction models [4].

Therefore, there comes a need to not just display the attributes of the river, but also reveal the topology of the river network system. To fulfill this requirement, I developed a Shiny application in R language with the function to visualize different pollutant of different measurement in river network and the variation of branching ratio in heat maps. Through this integrated Shiny program, users can select the interested river network system to view the specific river network properties and understand the topology logic of the corresponding river network.

## River Network Structural Properties and Branching Ratio

Many scholars have studied the structural properties of river networks and proposed theories such as self-similarity, random and random self-similarity. For example, Horton [5] and Strahler [6] graded the river reaches in the river network by the Horton-Strahler method. Statistics show that the relationship between reach number, length and corresponding catchment area and reach level conforms to Horton-Strahler's rule respectively, and structural parameters such as branching ratio, length ratio and area ratio are proposed. Based on this classification method, Tokunaga [7] introduced the concept of collateral ratio and defined the Tokunaga rule characterizing the relationship between river reach and sink, showing the self-similarity of river network topology [8].

Although some knowledge has been obtained about the river network structure and its relationship with the watershed context, how to better visualize the side branch ratio of the river network structure has not been deeply explored

Watershed has a certain shape, such as plume, leaf and branch [9], and parameters such as length, width and surface area are often used to determine the characteristics of the surface. The watershed river network in addition can be macroscopically characterized by watershed geometric parameters. In addition, the river level, branch ratio, length ratio, and area ratio can also be obtained. The geometric structure of the river network is deeply described. If we do not focus on geometric quantities such as river length and control area, and only focus on the number of rivers and the inflow and sink relationship between rivers, we can explore the topology of river networks.

The structural characteristics of river network are often expressed by Horton-Strahler method. After classify the river according to the Horton-Strahler classification method, for a river network with *N* levels, the number and the average length of rivers with level w can be statistically obtained, where w = 1,2,... , *N*. For a nature river network, Horton-Strahler's rule of river mesh and average length ­ is as follows :

(1)

(2)

Where is the branching ratio of river network, and its value is usually between 3 and 5. is the river network length ratio, which usually takes a value between 1.5 and 3.

Tokunaga method is often used to describe the topological structure characteristics of river networks. For a river network with Horton-Strahler level number *N*, the level 1 tributaries that sink into the level 1 tributaries are denoted as (1,1), and the number of level 1 tributaries with this type is denoted as . The level 1 branch that sinks into the level 2 branch is denoted as (1, 2), and the number of all such level 1 branches is denoted as . By analogy, the level *i (i≤j≤N)* tributaries that enter the level j tributaries are denoted as (*i, j*), and the number of such level *i* tributaries is denoted as . Then an upper triangular moment matrix is constructed for all (*i≤j≤N*).

For a river with Horton-Strahler level *i*, the number is:

(3)

Define the side branching ratio (*i < j≤N*) as follows:

(4)

If (where is a constant number), then the river network is topologically self-similar. This suggests that the river network has a similar structure at different scales, often associated with uniform topography and consistent erosion processes.

Figure 3 presents the upper triangular matrix of the side branching ratios for Waipaoa river. If the river network satisfies the topological self-similarity, the upper triangular matrix should satisfy (where is a constant number). However, we can observe from Figure 3 that , which shows the small degree variation of numerical value, and shows the leading function of self-similarity. While and shows large variation in values, which reveals reflects the influence of macroscopic conditions (such as geology) on the topology structure of the basin. Although the collateral ratio in the matrix decreases with the increase of the grade difference, the decreasing trend is more consistent, indicating that the river network maintains some self-similarity at different scales. This self-similarity may be related to the geographical and geological conditions of the river, reflecting the way the river network adapts to its environment.

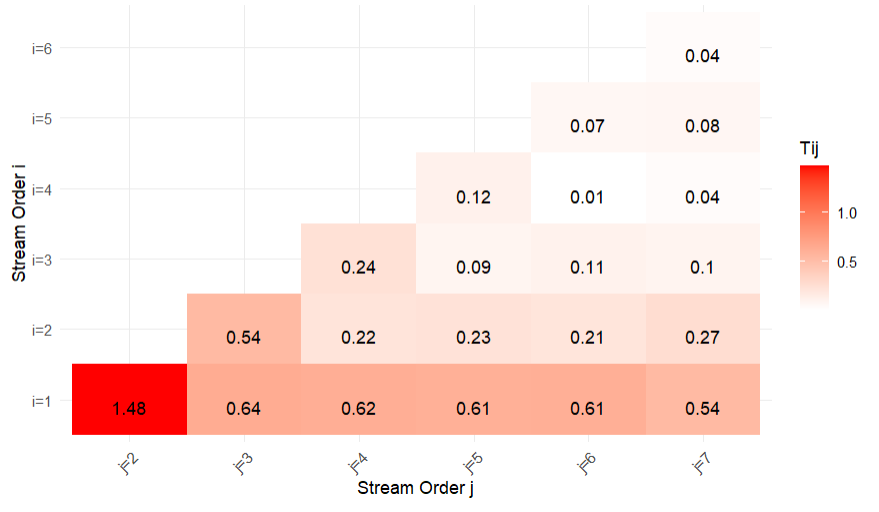


Figure 3. Side Branching Ratio for Waipaoa River

## Good vs. Bad Visualization Practices

Visualization plays a pivotal role in data analysis by turning complex data into comprehensible and actionable insights. However, the effectiveness of a visualization largely depends on how well it conveys the intended message and simplifies the complexity of the data.

Bad visualization usually have the following characteristics in common:

1.Complex and difficult to understand: If the viewer needs to spend a lot of time to understand the content of the visualization, then it is usually not a good design.

2.Information overload: Attempts to present too many data points or dimensions, resulting in visual clutter.

3.Using the wrong chart type: for example, pie charts for time-series data or bar charts for continuous variables.

4.Visual misinterpretation: such as the use of disproportionate graph sizes or inappropriate scales that distort the interpretation of the data.

Let’s take river visualization as an example. Figure 4 shows the river lines of Kahiaka river, and the thickness of the lines represents the difference in the Strahler grade of the river. The plot below is quite confusion without clear legend to show the Strahler level of the river lines, making it hard for reader to understand the message the diagram is trying to convey.

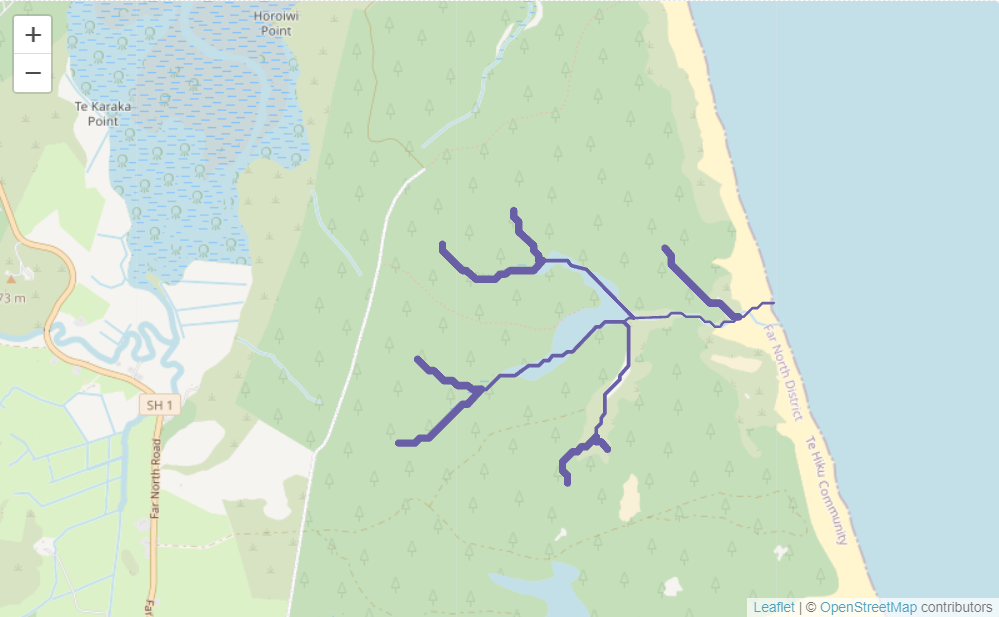


Figure 4 Example of Bad Visualization on Strahler Level of Kahiaka River

According to Rougier, Droettboom, and Bourne in their influential paper, "Ten Simple Rules for Better Figures" (2014)[10], effective data visualization plays a crucial role in communication and decision-making in scientific research. Building on the principles outlined by these authors, good visualization should incorporate several key characteristics to ensure both accuracy and clarity in presenting data.

These characteristics include:

1.Clarity: A good visualization should be intuitive, avoiding ambiguity and enabling the viewer to grasp the information quickly.

2.Relevance: Ensure that the presented data is directly related to the topic or problem being discussed.

3.Beautiful and simple: The design should be simple and attractive, using appropriate colors and graphics, but it should not be too loud to distract attention.

4.Accuracy: Present the data accurately and avoid misleading the audience.

Effective data visualisation is an essential component of the analytical process, not just a nice-to-have. Our approaches to data display must change along with the complexity of data and technology. Anyone working in data analysis and communication has to be aware of best practices for visualisation and always improving their methods and abilities. Clear glimpses or in-depth explorations of datasets supporting policies, innovations, and scientific discoveries are provided by well-designed visualisations, which are indispensable instruments for decision-making. These visual storytelling help make difficult knowledge understandable and useful.

According the rules above Figure 5 is generated to convey the same message as Figure 4 did. Figure 5 has a clear legend to make readers understand the message the plot is trying to convey and uses different colors and numbers to label river of different Strahler levels instead of using thickness to present the Strahler levels.

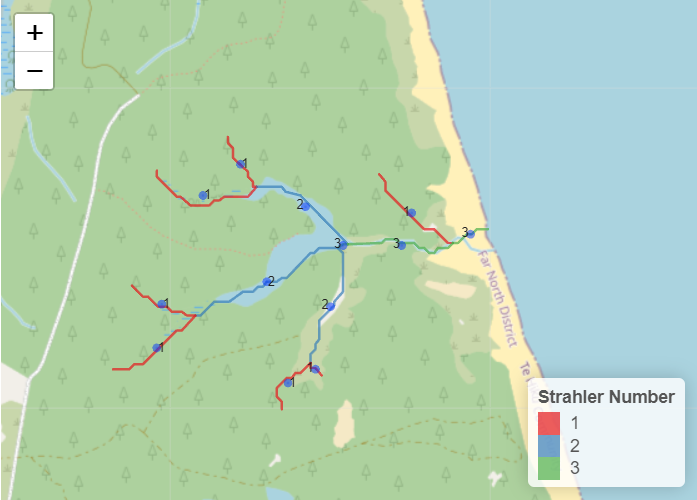


Figure 5. Example of Good Visualization on Strahler Level of Kahiaka River

Therefore, with proper visualization techniques, the same river lines can convey very different information. During the process I was designing the the visualization framework, clarity、simple and accuracy are always the primary goal. For this reason, I chose to create interactive clickable maps of the values and combine different shades of color to show the level of pollution[Figure 6].

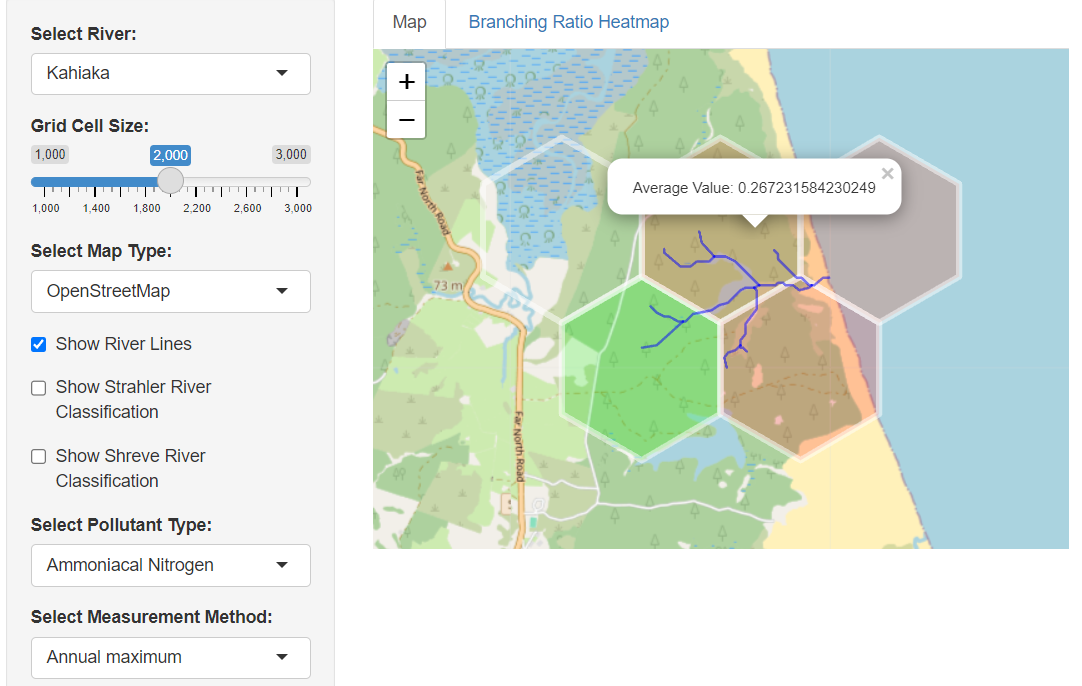


Figure 6 Example of River Visualization Project

## Thesis Outline

In this thesis, a tool is created to allow users to select Ammoniacal Nitrogen and Nitrate-Nitrite Nitrogen of two different measurement method to show values on maps. The R Shiny package is used to create a web-based tool.

In Chapter 2, we explore the functionalities of `Mapview` and Optimal Channel Network in R. Additionally, we will analyze how these two visualizations can contribute to the development of our final visualization design.

In Chapter 3, we discuss the Shiny package used to create the tool. The chapter is mainly focused on how we designed the Shiny app. Another focus is on explaining what insight can this kind of visualization offer.

In Chapter 4, we present our experimental design to evaluate the connectivity, geometry structure and topology structure within New Zealand rivers. We will also introduce the datasets we used and the area we studied.

In Chapter 5, we discuss the results of our experiment,

# Chapter 2 Related Work

This chapter discussed the techniques used for visualizing river lines. To be more specific, we will look into how we can implement river visualization in R. In Section 2.1 we introduce the two main ways of river visualization and the packages to implement them in R programming. In Section 2.2 we explore the `Mapview` package and we will look into how it presents the data. In Section 2.3 we discuss Optimal Channel Network's data acquisition in detail, as well as its set of visualization capabilities. In Section 2.4 we discuss what insight these two techniques offer. In Section 2.5 we explain what problems we should avoid when designing backend data processing method and how we preprocess the river data.

## 2.1 Existing Visualization Methods

At present, the two main ways of river visualization are primarily realized through two different types of data: one is through Shapefile datasets, and the other is through Digital Elevation Model (DEM) datasets. Shapefiles provide detailed geographical information, allowing for straightforward visualization of river networks and their attributes. In contrast, DEM datasets offer elevation data, enabling the creation of topographical visualizations that illustrate the flow dynamics and terrain features of river systems.

In the context of R programming, these visualization methods can be effectively implemented using different packages. The first method involves using Shapefiles in combination with the `Mapview` or `Leaflet` package, which facilitates the interactive mapping and visualization of geographical data. The second method leverages DEM datasets with the `OCNet` package to create detailed visualizations of river networks, incorporating elevation data to enhance the understanding of the river's topography and flow dynamics.

## 2.2 Mapview

The `Mapview` package provide a simple and interactive way to visualize spatial data, particularly useful for quick and straightforward geographic data representation. It supports a wide range of spatial objects and formats, including Shapefiles, and allows for the creation of interactive maps with minimal code.

Advantages of Using `Mapview`:

1.Ease of Use: `Mapview` is user-friendly and requires minimal coding effort to generate interactive maps, making it accessible for users with varying levels of programming expertise. For example, with just a few lines of code, users can load a Shapefile and visualize it interactively. The following code can be used to visualize spatial data:

river\_data <- st\_read("path\_to\_shapefile.shp")

mapview(river\_data)

2.Interactivity: Maps created with `Mapview` are highly interactive, allowing users to zoom, pan, and click on features to obtain more information. This interactivity enhances the user experience and aids in better data exploration. In terms of river spatial data, we could just see the data of each segment by clicking the corresponding line on the map. Figure 7 presents a visual example of creating the Marokopa river using the `Mapview` package.

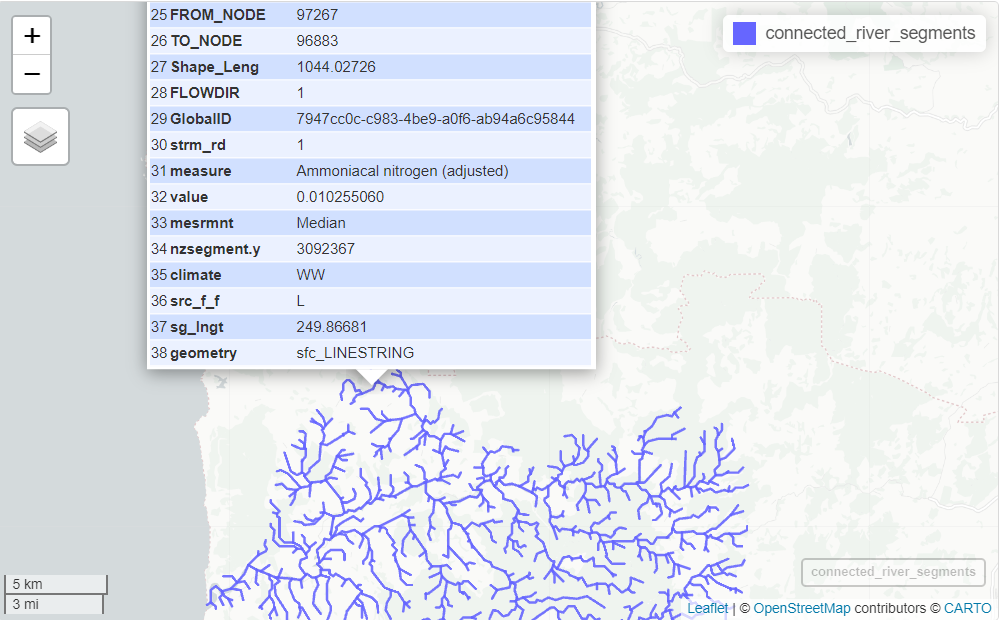


Figure 7 Example of `Mapview` interactivity

Disadvantages of Using Mapview:

1.Limited Advanced Features: Compared to other packages like leaflet, `Mapview` lacks some advanced customization options and functionalities. For instance, it may not support all types of interactive widgets or complex map layers. Users requiring highly customized map interactions might find `Mapview`'s capabilities insufficient.

2.Performance with Large Datasets: `Mapview` can struggle with performance issues when handling very large spatial datasets, leading to slow rendering times or even crashes. Visualizing a very large river network might be slow and less responsive. In terms of New Zealand rivers, there are 593517 segments rivers in total, which will cost huge amount of time of process and yet it will be hard to see each `tiny` segment in the whole map[Figure 8].



Figure 8 All river lines of New Zealand

## 2.3 Optimal Channel Network

The Optimal Channel Network (OCN) is a sophisticated tool for visualizing river networks using Digital Elevation Model (DEM) data. OCN utilizes optimization algorithms to simulate the natural formation and evolution of river networks. Specifically, it employs principles from optimal transport theory and network theory to determine the most efficient pathways for water flow across a landscape [1]. It is designed to accurately represent the physical layout of river systems in three-dimensional space, incorporating elevation and flow dynamics.

Advantages of Using OCN:

1. Accurate Representation: OCN provides a detailed representation of river networks by integrating elevation data, which facilitates comprehensive visualization of river topography and flow dynamics, in contrast to `Mapview`. Figure 9 depicts the Marokopa river as visualized by OCN. In comparison to Figure 7(visualized by `Mapview`), OCN's representation of the river is more accurate in restoring its original width.

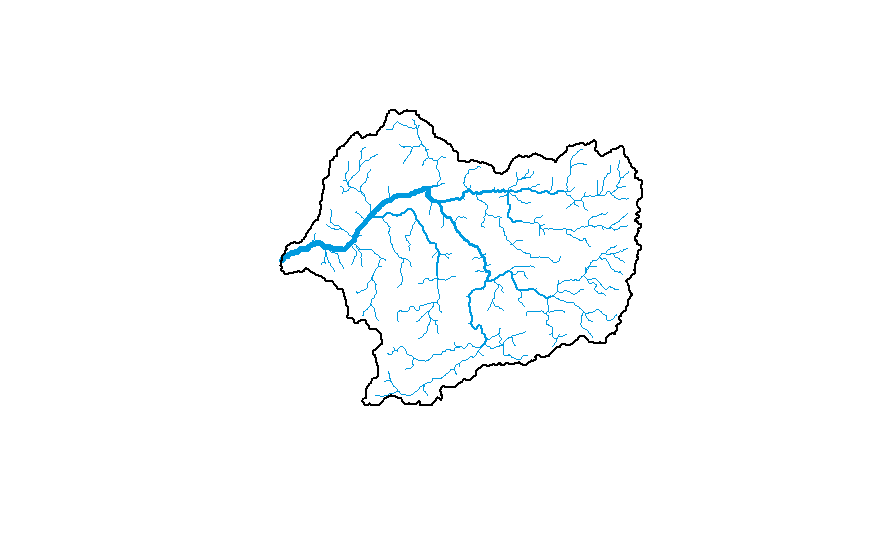


Figure 9 Example of Marokopa river Using OCN

1. Comprehensive Analysis: The tool supports detailed analysis of hydrological processes and river network structures, aiding in the study of water flow, sediment transport, and other related phenomena. OCN offers a range of aiding analysis tools, such as watershed division and meta-population modeling [Figure 10].

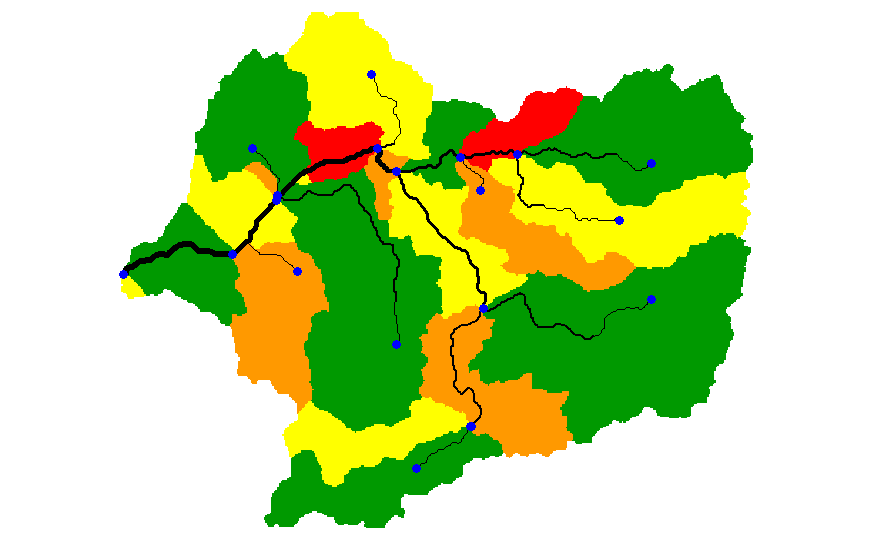


Figure 10 a. Subcathment of Marokopa river using OCN

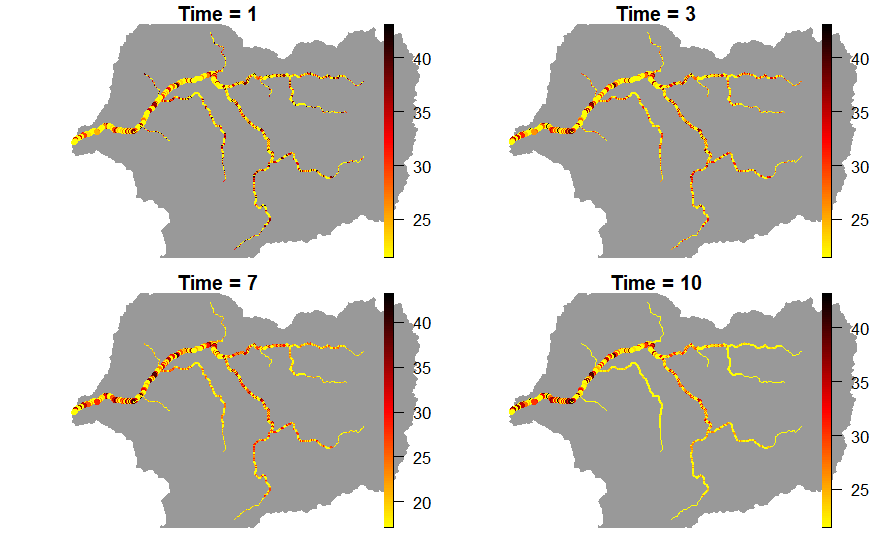


Figure 10 b. Metapopulation Simulation of Marokopa river using OCN

Disadvantages of Using OCN:

1.Data Requirements: The need for high-resolution DEM data is a limitation. Acquiring DEM data from R language requires the user to know the WGS84 geographic coordinates of the desired river outlet, and calculate the formula(5) according to the latitude and longitude to get a reasonable map resolution.

(5)

1. Computational Intensity:The optimization algorithms used by OCN can be computationally intensive, particularly for large datasets or complex simulations. For instance, to compute the subcathment of Marokopa river(medium sized river) would take 10 minutes for the aggregation process.

## 2.4 Insights from Visualization Techniques

The use of Mapview and the Optimal Channel Network (OCN) offers several important insights into river visualization techniques. Mapview highlights the significance of user-friendly interfaces and interactive data exploration. Its ease of use and dynamic capabilities make it accessible for broader audiences, facilitating preliminary data analysis. However, its limitations in advanced customization and performance with large datasets, such as the extensive river networks in New Zealand, underscore the need for more powerful tools.

OCN demonstrates the value of incorporating elevation data for accurate river network visualization. By using Digital Elevation Model (DEM) data, OCN provides a realistic representation of river topography and flow dynamics, which is essential for detailed hydrological analyses. The comprehensive analytical capabilities of OCN, including watershed division and meta-population modeling, highlight the benefits of integrating multiple functions within a single tool. However, OCN's high data requirements and computational intensity present challenges, particularly for large or complex river systems.

Taking the advantages of both techniques together, we should pay attention to the following points when designing visualizations:

·User Engagement and Accessibility: Simplicity and ease of use would make it highly accessible to users with varying levels of technical expertise.

·Interactivity: The ability to dynamically interact with the map, such as zooming, panning, and clicking on features to reveal more information, enhances the user’s ability to understand and analyze spatial relationships and patterns within the river network.

·Short response time: Users need to receive feedback quickly to maintain their engagement and productivity.

## 2.5 Enhancing Backend Processing Efficiency

When dealing with spatial data, one major problem is the long processing time. Both `Mapview` and OCN would take quite a long time to present a large river network. Especially for OCN, things like river extraction and data preprocessing can take at least ten minutes before performing the steps to display the river in 3 dimensional space.

To address this issue, we implemented a method to speed up backend processing. Our approach involves preprocessing the data by extracting the river network based on river HydroIDs and saving all related IDs of a river network into a text file. This way, the Shiny application only needs to read the IDs from the text file to extract the entire river segment, significantly reducing the processing time required to visualize the river network. The following code is used to extract river from the REC datasets:

findConnected <- function(targetID, riverLines, visited = numeric()) {

  queue <- c(targetID)

  while (length(queue) > 0) {

    currentID <- queue[1]

    queue <- queue[-1]

    if(currentID == -1) {

      next

    }

    if (!currentID %in% visited) {

      visited <- c(visited, currentID)

      downstreamIDs <- riverLines$HydroID[riverLines$NextDownID == currentID]

      upstreamIDs <- riverLines$NextDownID[riverLines$HydroID == currentID]

      for (id in c(downstreamIDs, upstreamIDs)) {

        if (!id %in% visited ) {

          queue <- c(queue, id)

        }

      }

    }

  }

  visited <- visited[visited != -1]

  return(visited)

}

The function inputs three parameters: the target river segment ID (`targetID`), the dataset containing river lines (`riverLines`), and an optional vector of already visited IDs (`visited`). The function initializes a queue with the targetID and iteratively explores connected segments. Within the loop, the function dequeues the first ID and checks if it has been visited. If not, it adds the ID to the `visited` list and finds all downstream and upstream IDs connected to the current segment. These connected IDs are then added to the queue for further exploration, provided they haven't been visited yet. The function ensures that IDs equal to -1 are excluded from the results (ID `-1` means the segment is the outlet of this river network), and returns a vector of all connected segment IDs. Then we store all the IDs for Shiny app.

# Chapter 3 Shiny App

While both the `Mapview` package and the Optimal Channel Network (OCN) tool offer valuable capabilities for visualizing river networks, they also present certain limitations that can impede comprehensive data analysis and visualization.

To overcome these limitations, a custom Shiny application is developed and a new approach for visualizing pollutant level is proposed. Shiny, a web application framework for R, offers an interactive and highly customizable platform for visualizing river networks. To meet the needs of regional pollutant value viewing, we designed adjustable size hexagonal areas for users to view the output values.

In this chapter, we will look into how we can implement river visualization in R and how to display the river related data and topology structure to end-users more efficiently and intuitively through R Shiny.The chapter is organised as follows. In Section 3.1 we introduce the input of Shiny programme and the function of our tool. In Section 3.2 we explain how to process the original data. Section 3.3 provides an overview of the backend. Section 3.4 provides an overview of the web-based Shiny application, detailing its layout, user prompts, input selection, and the visualization and data output for topology.

## 3.1 Shiny Function

The Shiny app is adapted to the data of 13 major rivers in New Zealand. The The New Zealand River Environment Classification V2.0 obtained from NIWA and the River Water Quality obtained from Stats NZ are used for the back-end datasets. The Leaflet package will be used to assist in the display of the map background and the river lines. The Leaflet package, created for interactive mapping, contains functions to display and customize map backgrounds, add various geographical features such as river lines, and enable intuitive user interactions, making it ideal for visualizing complex spatial data.

The Shiny app offers these following functions:

|  |  |
| --- | --- |
| **Function Name** | **Function Description** |
| River Selection | Users can select the river they want to check |
| Grid Size Selection | Users can set the grid size of the hexagon from 1000 to 3000 |
| Map Type Selection | Users can choose from three map background styles |
| Show River Lines | Click to display river lines on the map |
| Show Strahler River Classification | Click to display Strahler Classification of each river segment on the map |
| Show Shreve River Classification | Click to display Shreve Classification of each river segment on the map |
| Select Pollutant Type | Select Ammoniacal Nitrogen(AN) or Nitrate-Nitrite Nitrogen (NN)from the drop-down menu |
| Select Measurement Method | Select two different measurement for the corresponding pollutant from the drop-down menu |
| Branching Ration Heat Map | Click to view the branching ration heat map for the selected river |

## 3.2 Pre-processing

The type of data we used is `Shapefile` from NIWA and Stats NZ. Shapefile is a widely-used geospatial vector data format for geographic information system (GIS) software. Developed by Esri [11], it stores geometric location and associated attribute information of geographic features.

In terms of R programming, `Shapefile` is typically read in and processed with `sf` or `sp` package. Since the rivers we need to process are large, and `sf` is more suitable for processing large spatial data than `sp`, we chose `sf` to read and process the data. The `sf` package supports a wide range of spatial data formats and integrates seamlessly with other R packages for data visualization and analysis [12].

The datasets used in this study is derived from the River Environment Classification (REC) database and Stats NZ. The two datasets are joined together using the `nzsegment` keyword, which represents the unique id of each river in New Zealand. There are 44 attributes of river in total in these two datasets, we first filtered out the 9 columns attributes we need, which are listed as follows:

|  |  |
| --- | --- |
| **Attribues** | **Description** |
| HydroID | Sequential numbers from 1 to the maximum number in the dataset, representing the ID of this hydro segment. |
| NextdownID | Represents the `HydroID` of the downstream hydro segment. |
| StreamOrde | The Strahler level of this segment. |
| Measure | The type of pollutants present in the river reach. |
| value | The measured values of the pollutants. |
| mesrmnt | The methods used for measuring the pollutants. |
| Shape\_Leng, | The length of the river reach. |
| CUM\_AREA | The cumulative watershed area in square meters (m²). |
| geometry | Geolocation information in the form of `LINESTRING` type. |

The next step in our process is to convert the geolocation coordinates of our spatial data to ensure compatibility with the visualization tools we are using. Specifically, the `leaflet` package only supports the WGS84 coordinate system (EPSG:4326). Therefore, it is necessary to transform the coordinates of our `sf` object to this coordinate system.

Since the data stores information about all the rivers in New Zealand, it would take a long time to load. Therefore, we choose to use the `findConnect` function (mentioned in Section 2.5) to extract the ids of 13 large river networks in New Zealand during data pre-processing.

## 3.3 Backend

All the formula calculation, processing of river lines and graphics is done in the background of R program. An end-user of the Shiny app would select the desired river, pollutant parameters(AN and NN) and measurement(Annual maximum/Median for AN and Median/95th for NN) to generate the map on the right side of the user interface.

The data set utilized in the application comprises stream order, geometric coordinates, measurement, hydro ID, and next downstream ID of 62578 river segments encompassing a total of 13 rivers. The code can be found here [Yishion1/New-Zealand-River-Visualization-based-on-Optimal-Channel-Network (github.com)](https://github.com/Yishion1/New-Zealand-River-Visualization-based-on-Optimal-Channel-Network) . The follow plots [Figure 11] are the output of the Shiny application.

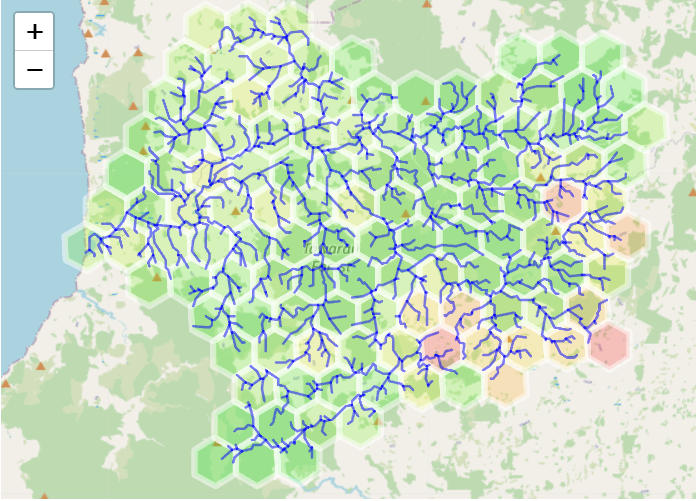


Figure 11a. River Visualization Shiny App River Lines Output



Figure 11b. River Visualization Shiny App Branching Ratio Output

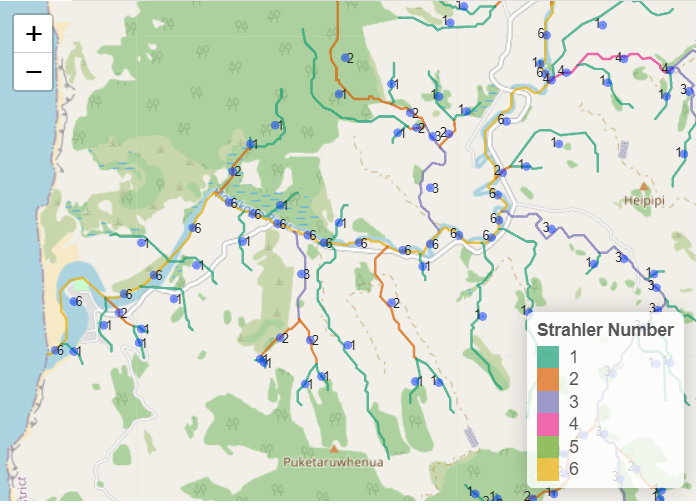


Figure 11c. River Visualization Shiny App Strahler Classification Output

The following code was used to create the hexagon visualizations in the Shiny:

reactiveHexRivers <- reactive({

    req(reactiveRiverData())

    progressInfo("Generating hex grid... ✓<br>")

    rivers\_sf <- reactiveRiverData()

    hex\_grid <- st\_make\_grid(rivers\_sf, cellsize = input$gridsize, square = FALSE, what = "polygons")

    hex\_sf <- st\_sf(geometry = hex\_grid)

    hex\_rivers <- st\_join(hex\_sf, rivers\_sf, join = st\_intersects) %>%

      group\_by(geometry) %>%

      summarize(Value = mean(value, na.rm = TRUE), .groups = 'drop') %>%

      st\_transform(crs = 4326)

    hex\_rivers <- hex\_rivers[!is.na(hex\_rivers$Value), ]

    progressInfo("Hex rivers data prepared... ✓<br>")

    list(hex\_rivers = hex\_rivers, rivers\_sf = rivers\_sf)

  })

The function is a reactive function in a Shiny application that processes spatial river data to create a hexagonal grid visualization. It starts by ensuring the availability of river data through the `req` statement. The function then indicates the progress of generating the hex grid with `progressInfo`. It loads the river data into `rivers\_sf` and creates a hexagonal grid over the extent of the river data using `st\_make\_grid`, with the cell size determined by the user-defined `input$gridsize`. The hexagonal grid is converted into an `sf` object, `hex\_sf`. A spatial join between `hex\_sf` and `rivers\_sf` is performed using `st\_join`, and the data is grouped by the hexagonal geometries. The mean value of the river data within each hexagon is calculated using summarize, and the resulting hexagons are transformed to the EPSG 4326 coordinate reference system with `st\_transform`. The function filters out hexagons with `NA` values in the `Value` column to remove those that do not contain river data. The progress is updated with another `progressInfo` call, and the function returns a list containing the filtered hexagonal grid (`hex\_rivers`) and the original river data (`rivers\_sf`). This process ensures that the final visualization includes only relevant hexagons. The ``rivers\_sf` here would later be taken as the input to draw river lines on the map.

## 3.4 Methods

The final app is a web-based application and can be found here:[Yishion1/New-Zealand-River-Visualization-based-on-Optimal-Channel-Network (github.com)](https://github.com/Yishion1/New-Zealand-River-Visualization-based-on-Optimal-Channel-Network) .

This system allows users to adjust the size of the hexagonal grid cells and view pollutant concentrations across different segments of the river. The design of this interactive feature provides several key benefits for environmental analysis and decision-making.

The dynamic hexagonal grid system is implemented using the `leaflet` package in R, which supports interactive mapping and spatial data visualization. The hexagonal grids are generated based on user-specified cell sizes, providing a flexible and detailed view of pollutant distributions. The core idea behind this design is to offer an intuitive and adjustable visualization tool that enhances the user's ability to analyze spatial patterns and trends in water quality data.

The landing page of our Shiny application features a sidebar for user input and two-panel tabs. The sidebar remains fixed, allowing users to switch between the tabs without losing their input settings. This layout ensures that users can seamlessly adjust their input values while navigating different views within the application. The straightforward design makes it clear where inputs should be entered, enhancing user experience and ease of use. The sidebar provides various input controls, including dropdown menus and sliders, to customize the visualization. Users can choose the river of interest from a dropdown menu, adjust the size of the hexagonal grid cells using a slider input, select the base map type (e.g., OpenStreetMap, Esri World Image, CartoDB Positron) from a dropdown menu, and toggle the display of river lines, Strahler river classification, and Shreve river classification using checkboxes. Additionally, users can select the type of pollutant to visualize and choose the measurement method for the selected pollutant.[Figure 12]

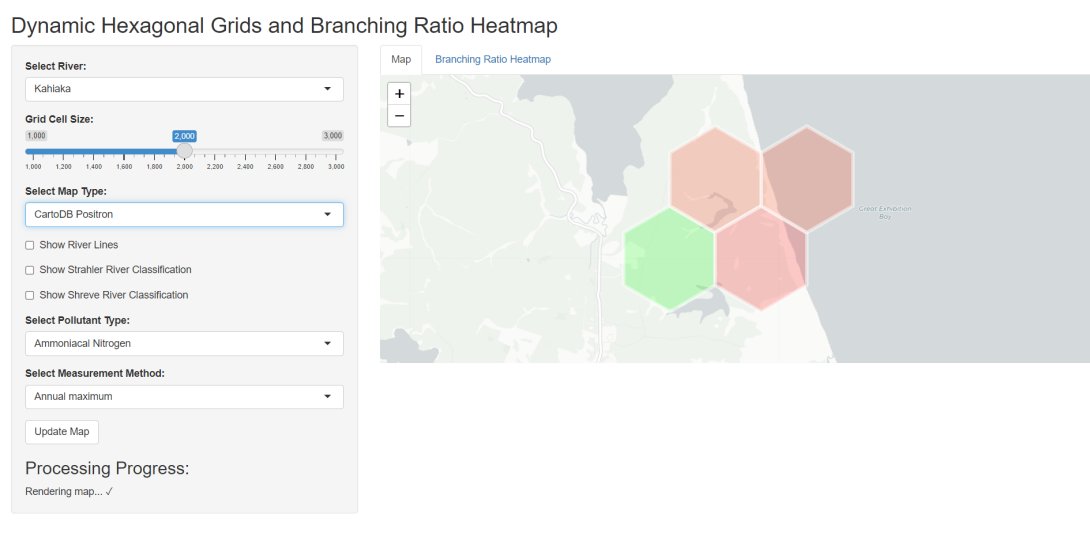


Figure 12. Landing Page of Shiny App

The main panel of the application consists of two tabs. The first tab, "Map," displays an interactive map with hexagonal grids and river lines. Users can interact with the map by zooming, panning, and clicking on hexagonal cells to view detailed pollutant information. [Figure 13]

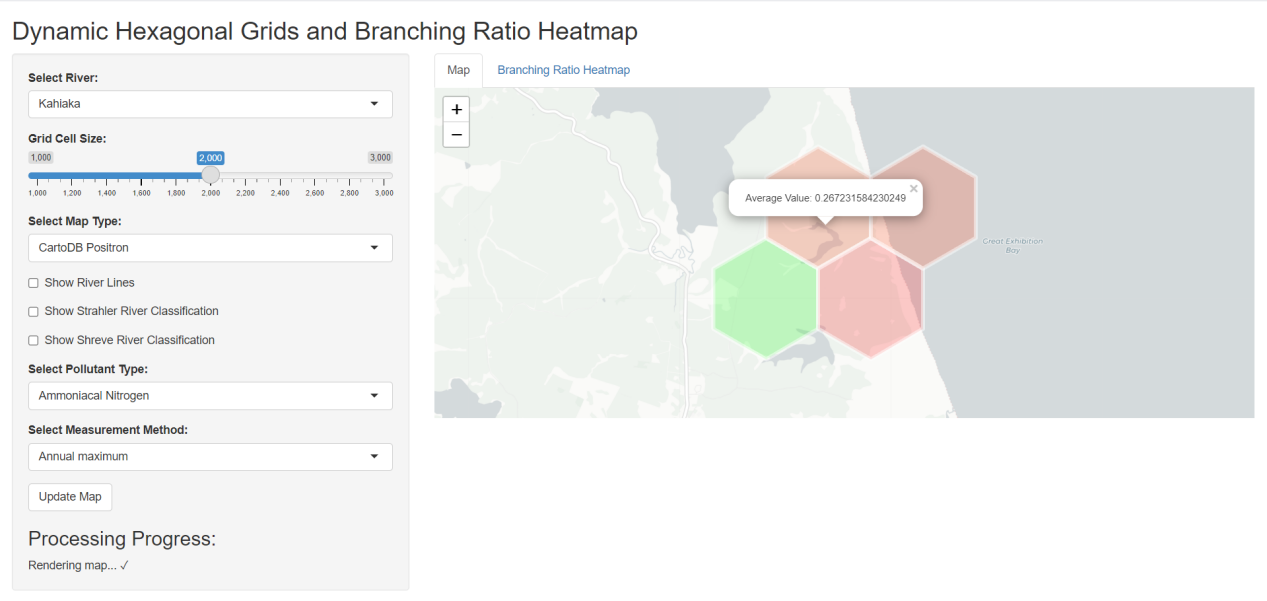


Figure 13 Shiny App Displaying The Value

The grid cells dynamically update based on the selected cell size and pollutant type.[Figure 14]

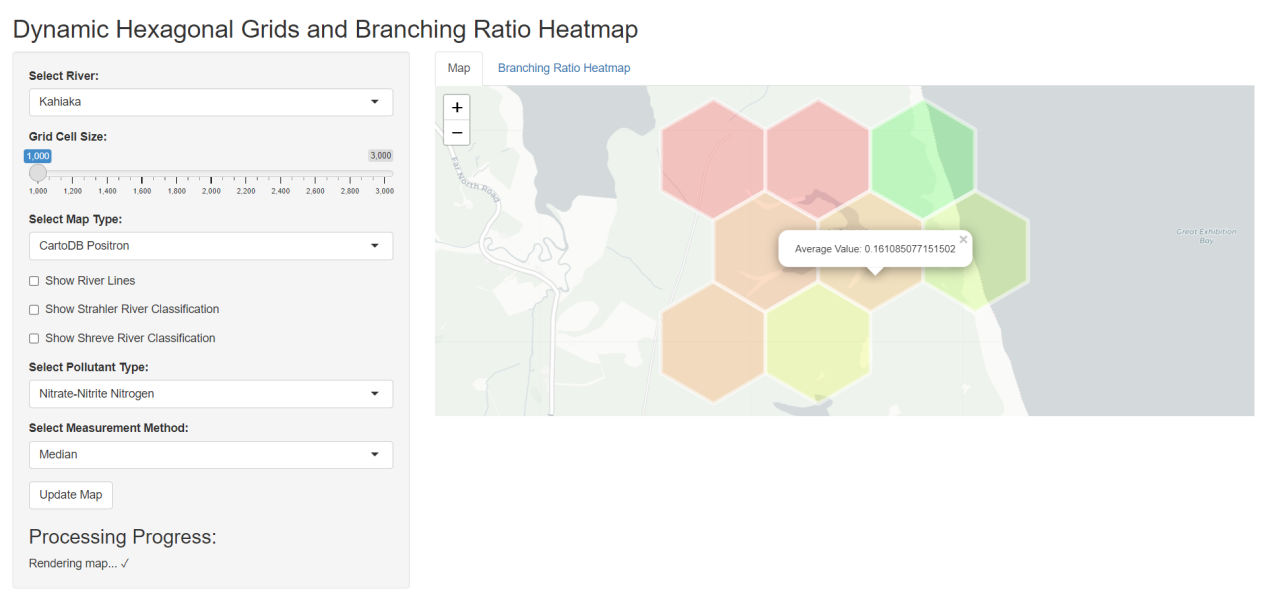


Figure 13. Shiny App after changing the grid size and pollutant type

By selecting the "Show River Lines" option, users can view river lines on the map on the right. The reason for designing this feature, instead of displaying river lines directly on the map, is that for large river networks, displaying all the river lines simultaneously can lead to visual clutter. Additionally, users might sometimes be more interested in regional data rather than the specific flow direction of the rivers.

The second tab, "Branching Ratio Heatmap," provides a heatmap visualization of the branching ratios within the river network. This heatmap helps users understand the topological structure of the river network and identify areas with varying branching ratios. [Figure 14]



Figure 14. The Branching Ratio Heat Map of Shiny App

The design and functionality of our Shiny application's user interface provide an intuitive and powerful tool for visualizing and analyzing river pollution data. By offering interactive features and customizable options, the application enhances the user's ability to explore spatial patterns and make informed decisions regarding environmental management. Meanwhile the branching ratio heat map reveals insight into river topology logic and connectivity. It is essential to discuss the connectivity within these river networks.

# Chapter 4 Experimental Design

In this chapter, we present our experimental design to evaluate the connectivity and topology structure within New Zealand rivers. The chapter is organised as follows. In Sections 4.1 we define the methods used to study the river connectivity. Section 4.2 describes the datasets as well used in river connectivity experiments.We present river topology analysis method in Section 4.3. Section 4.4 and 4.5 discuss the setting for river topology experiment. Section 4.6 summarize the setting for these two experiments.

## 4.1 Methods for Connectivity

The methods used to study river connectivity are defined in this section. The connectivity of river networks is crucial for understanding the movement of water, sediment and organisms within these systems. Connectivity influences several ecological and hydrological processes, including nutrient cycling, habitat connectivity and pollutant dispersal (Benda et al., 2004; Pringle, 2001)[13].

We define the influence of each river node in the following formula:

**Definitions:**

**Influences Initialization:** Initialize influence values for all nodes.

In this experiment, we further explored the impact of river connectivity on the distribution of river segment influence. The specific process is as follows:

**1.Source Node Identification:**

Similar to the previous experiment, we first identified source nodes (i.e., river segments with no upstream nodes) to initialize the calculations. These source nodes served as the starting points for influence calculation.

**2.Traceback Matrix and Steps Matrix Construction:**

Using river connectivity data, we constructed a traceback matrix that recorded the upstream nodes for each node. Additionally, we calculated a steps matrix to determine the number of steps from each node to its downstream nodes.

**3.Influence Initialization:**

We initialized the influence of each river segment to 1 and assigned a unique identifier to each segment for reference in subsequent calculations.

**4.Dynamic Influence Calculation:**

We used a queue to process all source nodes and gradually expanded to the entire river network.

For each node, we calculated the influence from its immediate upstream nodes and accumulated this influence to the current node, applying an exponential decay factor based on the distance (upstream influence weight).

The processed nodes were marked as calculated, and their downstream nodes were added to the queue to continue the influence calculation until all nodes were processed.

**5.Connectivity Variation Experiment:**

To study the impact of different connectivity parameters on river segment influence, we varied the connectivity parameter from 1 to 20 (in steps of 3) and calculated the normalized influence for all segments in the river network for each connectivity value.

The influence results were normalized to ensure comparability across different connectivity values.

The normalized influence results for different connectivity values were collected and organized into a data frame.

**6.Results Visualization:**

We used density plots to display the normalized influence distribution for different connectivity values, providing a visual representation of how connectivity variation affects the distribution of river segment influence.

To analyse river connectivity, we used several functions to trace river segments, calculate steps to downstream nodes, identify source nodes and dynamically calculate influence weights. Below is a detailed explanation of the methods and functions used in this analysis.

## 4.1.1 Traceback Sequence Analysis

We developed the `get\_traceback\_REC` function. This function generates a traceback sequence for each river segment by identifying upstream paths based on a nearest neighbour connectivity distance. The function takes connectivity and our river data set——`river\_segments` as input. For each node in this segments, the algorithm try to find the nearest upstream neighbour that is connected to this node within the number of connectivity and return an upstream path matrix of each node as results.

Function get\_traceback\_REC(connectivity, river\_segments):

    Initialize traceback\_results as an empty list

    For each node in river\_segments:

        Set HydroID to the current node's ID

        Set current\_neighbors to upstream neighbors of HydroID

        Initialize path with HydroID

        For i from 1 to connectivity:

            If current\_neighbors is empty, break

            Calculate distances to all upstream neighbors

            Select the nearest neighbor

            Add the selected neighbor to the path

            Update current\_neighbors by removing the selected neighbor and adding its upstream neighbors

        Add path to traceback\_results

    Convert traceback\_results to a matrix

    Return traceback\_matrix

The detailed information of inputs for `get\_traceback\_REC` are listed as follows:

|  |  |
| --- | --- |
| **Name** | **Description** |
| connectivity | Defines how many upstream euclidean distance nearest neighbors are connected to it. |
| river\_segments | Datasets that store the information of the location, hydro id, and join order of each segment |

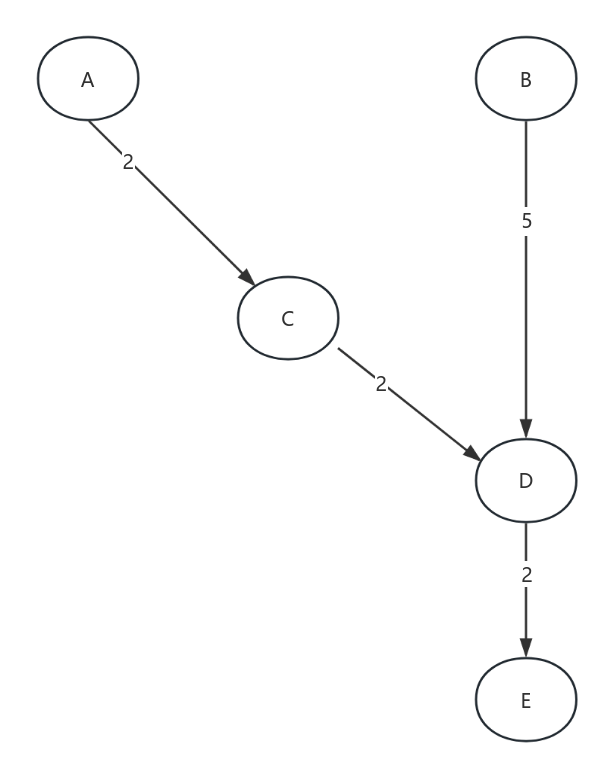


Figure 15. Example of Nearest Neighbour Connectivity

Let’s take Figure 15 as an example to explain the Nearest Neighbour Connectivity. The number on each edge represent the euclidean distance between two river node. The direction of the arrow represents the flow of the river. Now if we set connectivity to 2 and the above river network as input to `get\_traceback\_REC` and we wanna find the 2 nearest upstream neighbour of node E. Under this condition, node D is the first neighbour, then we compare the distance between node E and node C and the distance between Node E and node B. We find that node C has a shorter path to node E, thus node C is the second nearest neighbour of node E. The 2 nearest neighbour of node E is D and C.

## 4.1.2 Calculating Steps to Downstream Nodes

Then we define the ` calculate\_steps\_to\_downstream\_REC` function as below. The calculate\_steps\_to\_downstream\_REC function calculates the number of steps required for each node in the traceback path to reach its downstream node. This function helps in quantifying the influence of upstream segments on downstream segments.

Function calculate\_steps\_to\_downstream\_REC(traceback\_matrix, river\_segments):

    Initialize steps\_matrix with NA values

    Create a mapping of HydroID to NextDownID

    For each row in traceback\_matrix:

        Set steps\_matrix[i, 1] to 0 (source node)

        For each column in the current row:

            Initialize steps to 0

            Set current\_node to the current column value

            While current\_node is not NA and not the source node:

                If current\_node has a downstream node:

                    Update current\_node to the downstream node

                    Increment steps

                Else:

                    Set current\_node to NA and break

            If current\_node is the source node, set steps\_matrix[i, j] to steps

    Return steps\_matrix

## 4.1.3 Dynamic Influence Calculation

The `calculateInfluenceDynamically\_REC` function below calculates the influence of each river segment dynamically, considering both upstream influence weights and connectivity. This function iteratively updates influence values based on the cumulative impact of upstream segments.

Function calculateInfluenceDynamically\_REC(river\_segments, upstreamInfluenceWeight, connectivity, influences):

    Get source\_nodes using get\_source\_node\_REC function

    Get traceback\_matrix using get\_traceback\_REC function

    Get steps\_matrix using calculate\_steps\_to\_downstream\_REC function

    Initialize influences with default values

    Initialize nodes\_queue with source\_nodes

    Initialize calculated\_nodes as an empty list

    While nodes\_queue is not empty:

        Get the first node from nodes\_queue

        If current\_node is not in calculated\_nodes:

            Add current\_node to calculated\_nodes

            For each row in traceback\_matrix where the first column matches current\_node:

                For each column in the row:

                    Get upstream\_node and step

                    If upstream\_node is valid and step is greater than 0:

                        Update influences for current\_node

            Add downstream\_nodes of current\_node to nodes\_queue

    Return influences

## 4.2 Datasets

The datasets used in this experiment are derived from the River Environment Classification Version 2.0(REC2) database. The REC provides comprehensive information about New Zealand's rivers, including over 30 attributes for each river segment. These attributes cover various aspects such as hydrology, geology, straight line distance of a reach, and land cover. Real world datasets like REC2 allows us to run simulation on real world river networks. More importantly we could explore whether there is a pattern within these networks.

The relevant attributes used in REC2 are listed as follows:

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Type** | **Description** |
| HydroID | Numerical | Sequential numbers from 1 to the maximum number in the dataset, representing the ID of this hydro segment. |
| NextDownID | Numerical | Represents the `HydroID` of the downstream hydro segment. |
| UpcoordX | Numerical | Real Easting of the upstream end of a river segment in m (NZTM2000). |
| DowncoordX | Numerical | Real Easting of the downstream end of a river segment in m (NZTM2000). |
| UpcoordY | Numerical | Real Northing of the upstream end of a river segment in m (NZTM2000). |
| DowncoordY | Numerical | Real Northing of the downstream end of a river segment in m (NZTM2000). |

## 4.2.1 Area of study

We choose the Marokopa River as the study case. The Marokopa River is located on the west coast of New Zealand's North Island with geographical coordinates between approximately 38°15's and 38°20's and between 174° 45’e and 175° 00’e. The total length of the main stream is about 56 km, the width of the basin is about 30 km, and the drainage area is about 225 square km. The upper and middle reaches of the basin are more open, while the lower reaches are relatively narrow. The Marokopa River is characterized by its deep main flow and dense branches, and the river specific fall varies greatly, with the measured channel specific fall of about 10% over the years.

The soil in the basin is relatively loose and covered with a thick layer of volcanic ash soil. Due to the relatively concentrated rainfall in this area, the river is prone to form multiple tributaries and gullies during the wet season.

## 4.3 River Topology Experiment

As mentioned in Section 1.3, the structural characteristics of river network are often expressed by Horton-Strahler method. We can calculate the number and average length of rivers of level i according to (1)(2). Then we can use Tokunaga method to calculate the branching ratio of the river according to (3)(4).

## 4.4 Topology Study Area

We choose Marokopa River, Wairoa River, Whakatane River, Waipaoa River, Tukituki River as our study areas.

**1. Marokopa River**

The Marokopa River is located on the west coast of the North Island of New Zealand, with a total length of about 56 km and a drainage area of about 225 square km. The upper and middle reaches of the river are relatively open, while the lower reaches are relatively narrow, and the river specific fall varies greatly. The Marokopa River is characterized by its deep main stream and dense tributaries.

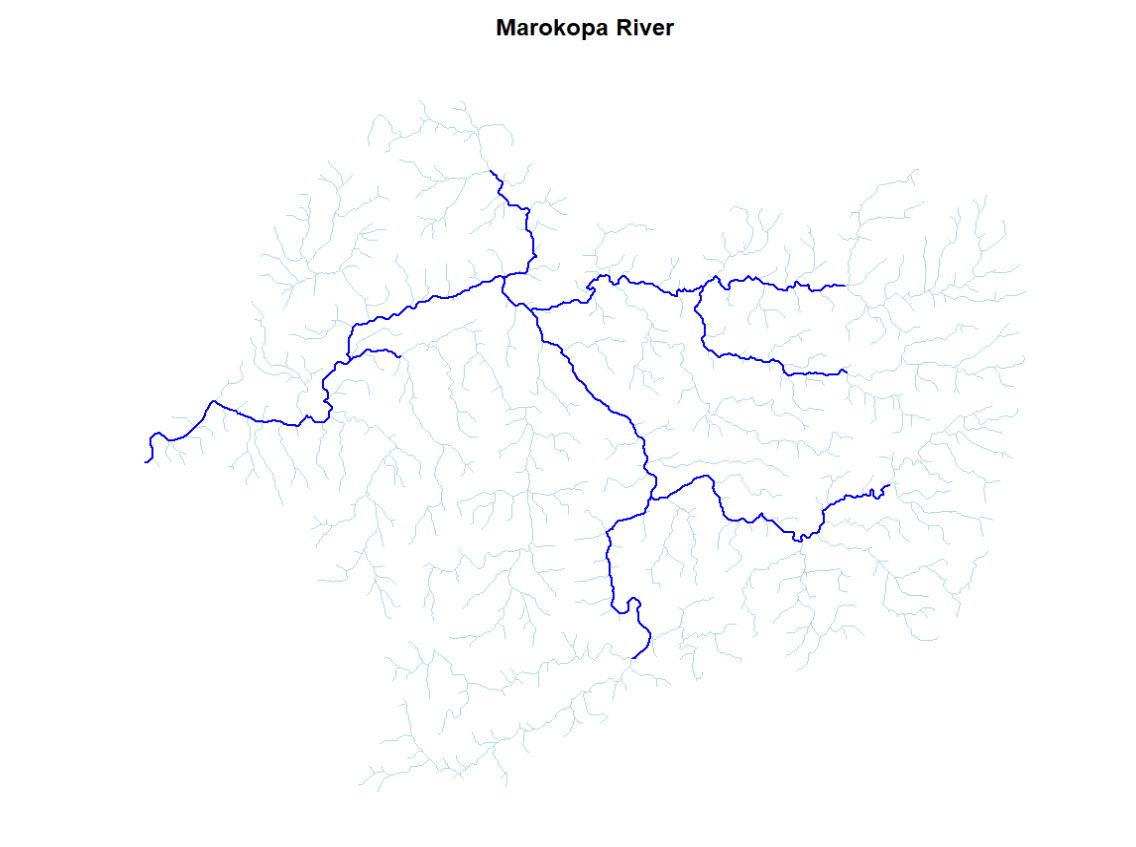


Figure 16. River Network of Marokopa River

**2.Wairoa River**

Situated on the eastern coast of New Zealand's North Island, the Wairoa River originates from Lake Ngamotu and traverses several small towns and agricultural areas before discharging into Hawke Bay. Spanning approximately 65 km in length with a drainage area of about 700 square km, the Wairoa River catchment encompasses a diverse range of geological features including mountains, hills, plains, extensive native forest, and agricultural land. These complex hydrological characteristics make it an ideal location for studying the impact of different terrains on river morphology.

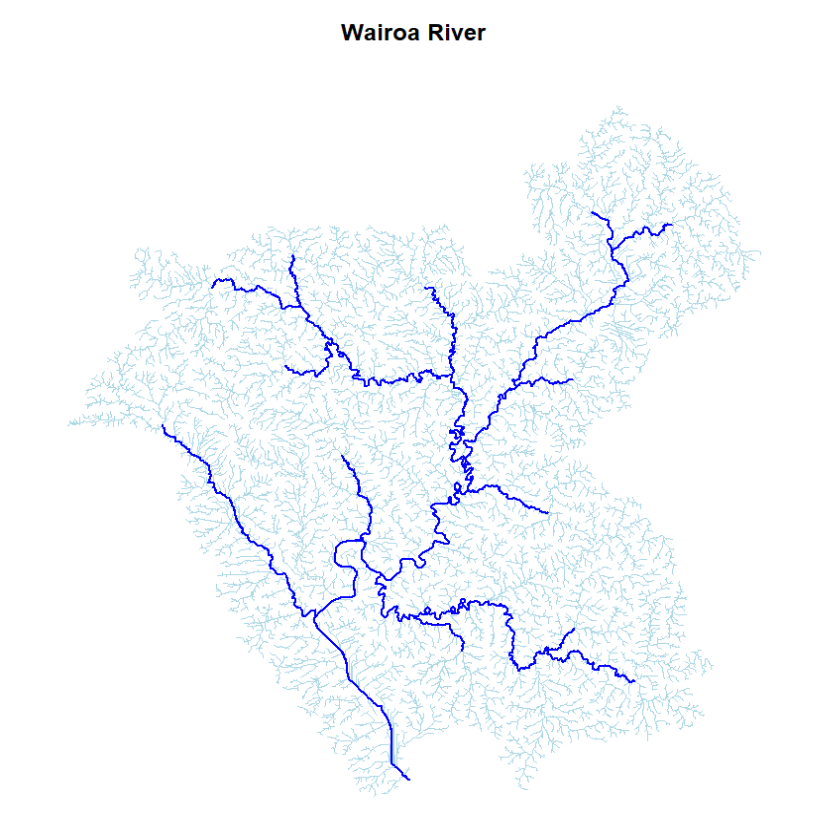


Figure 17. River Network of Wairoa River

**3. Whakatane River**

The Whakatane River, situated on the east coast of the North Island of New Zealand, originates from the Urewera Mountains. It spans approximately 95 km in length and covers a drainage area of about 960 square km. The river traverses through various urban and rural areas before emptying into the Bay of Plenty. The Whakatane River basin encompasses forested, agricultural, and urban regions, with the area it passes through characterized by high levels of human activity, particularly in agricultural and urban development.

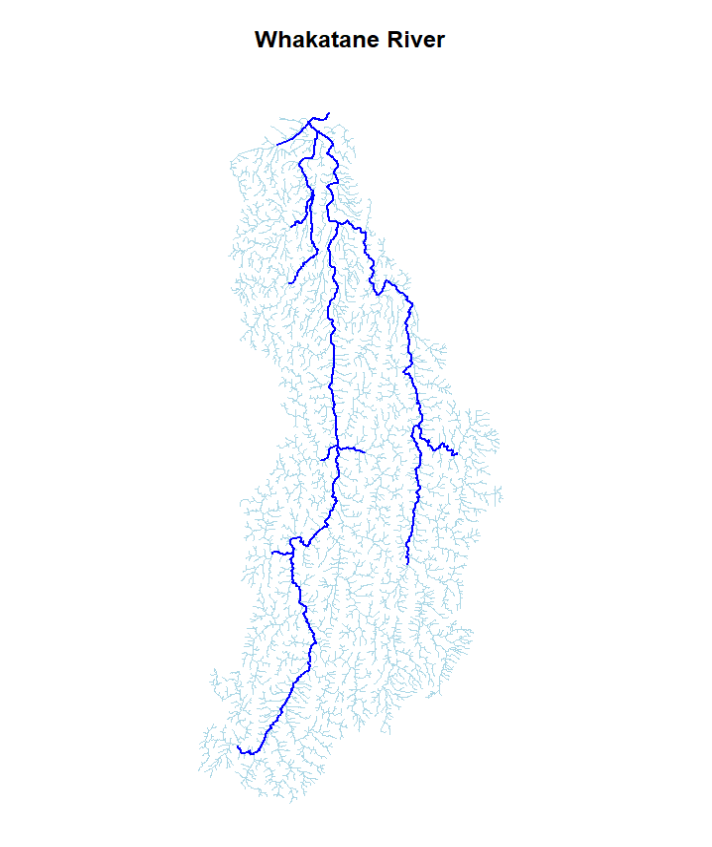


Figure 18. River Network of Whakatane River

**4. Waipaoa River**

The Waipaoa River is located on the east coast of New Zealand and originates from the Raukumara Mountains. It is about 130 km long and has a drainage area of about 2200 square km. The river flows through a vast agricultural area before emptying into Poverty Bay. The Waipaoa River watershed has significant sedimentary features with extensive farmland and grassland on both sides of the channel.

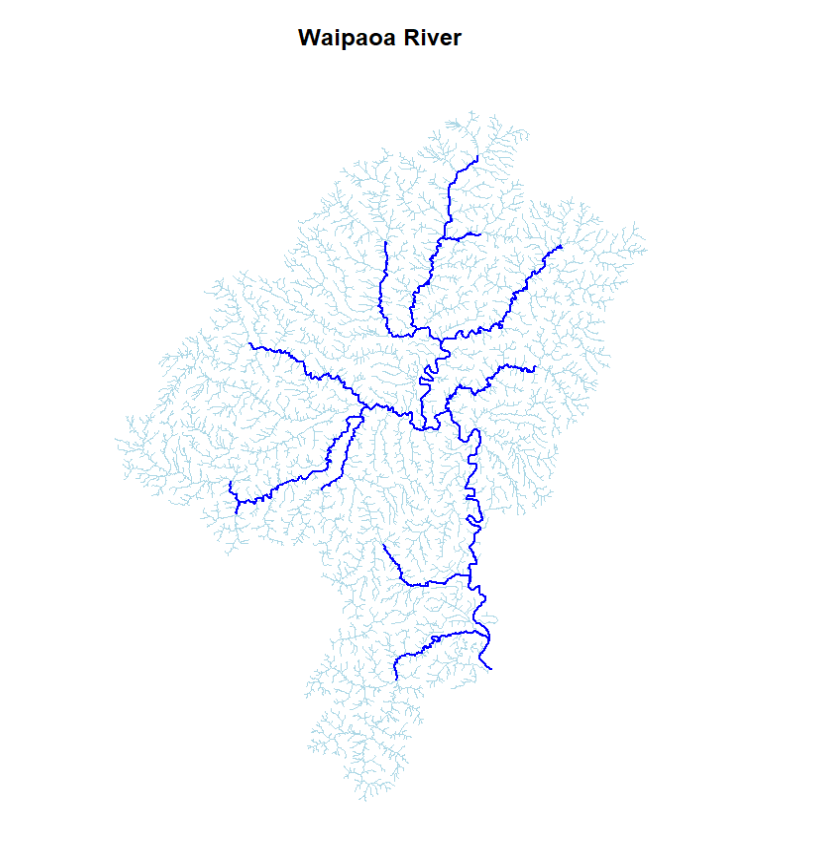


Figure 19. River Network of Waipaoa River

**5. Tukituki River**

The Tukituki River is located in the southeast of the North Island of New Zealand and originates from the Ruahine Mountains. It is about 117 km long and has a drainage area of about 2500 square km. The area through which the river flows is mostly mountainous and hilly terrain, with a complex channel network and numerous branches, and finally merges into Hawke Bay.

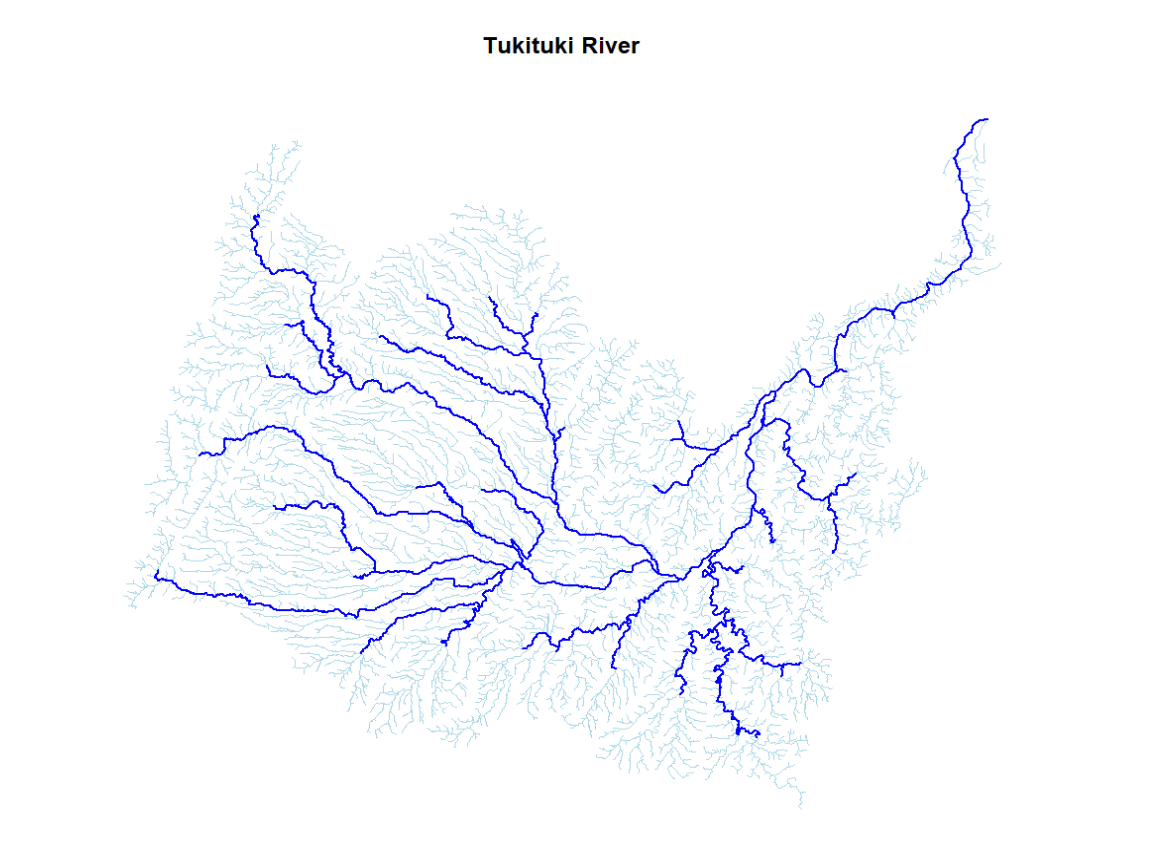


Figure 20. River Network of Tukituki River

These rivers were chosen for topological studies because of their diverse geological and geomorphic features. For example, the upper and middle reaches of the Marokopa River are relatively open, and the lower reaches are relatively narrow; the Wairoa River basin has a large amount of native forest and agricultural land; the Tukituki River flows through mountainous and hilly terrain, with a complex channel network and numerous branches.

## 4.5 Results of river network extraction

Figure 16-20 show what these five rivers look like, the plot below shows the relationship between the number of rivers and the number of river levels and the average length of river levels.

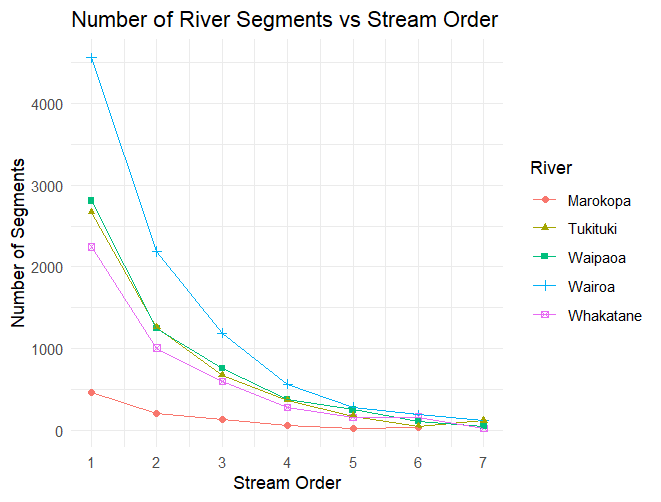


Figure 21 Relationship Between the Number of Rivers and the Number of River Levels

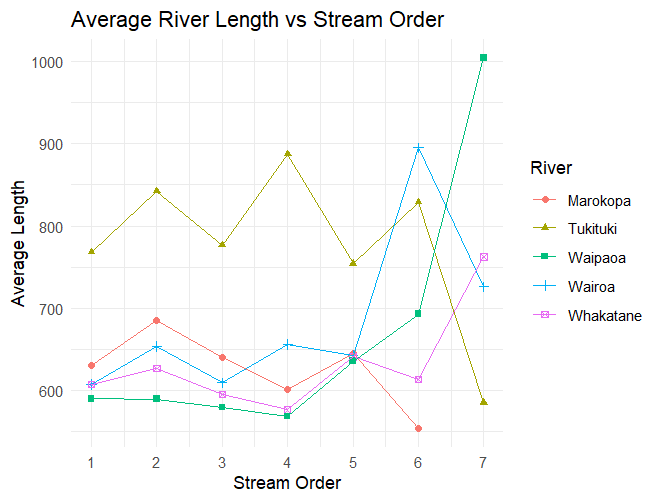


Figure 22. Relationship Between the Number of Rivers and the Average length of River Levels

From figure 16-20, we could see that the Marokopa and Waipaoa rivers exhibit dendritic network characteristics, with branches that are relatively dispersed and deeply incised. In contrast, the Wairoa and Whakatane rivers exhibit pinnate network characteristics, with dense and evenly distributed branches. The Tukituki River combines these two characteristics, showing a complex branching pattern.

From Figure 21, we could see the relation within number of rivers and its levels. It can be seen that the number of rivers in the five different basins are well distributed linearly at relatively low river classes (within class 2-4). However, at higher river class numbers (higher than 4), the data points have deviated from the linear distribution of higher class rivers. This suggests that the lower river classes satisfy the Horton-Strahlr law as shown in Eqs. (1) and (2) well and exhibit structural self-similarity; Lower order streams struggle to satisfy the Horton-Strahler rule, reflecting the fact that main streams and their major tributaries may exhibit consistent patterns and distributions.

## 4.5 Summary

In this chapter, we have presented our experimental design. We have introduced the datasets(REC2) and the selected attributes we used in our experiments. We then introduced the different areas(Marokopa River, Wairoa River, Whakatane River, Waipaoa River, Tukituki River) we studied in these two experiments. After that, we provide a brief overview of the results of river networks extraction.

# Chapter 5 Results

In this chapter, we present the results of our experiments on the REC2 datasets. The Chapter is organised as follows: In Section 5.1 we

## 5.1 Result for Connectivity

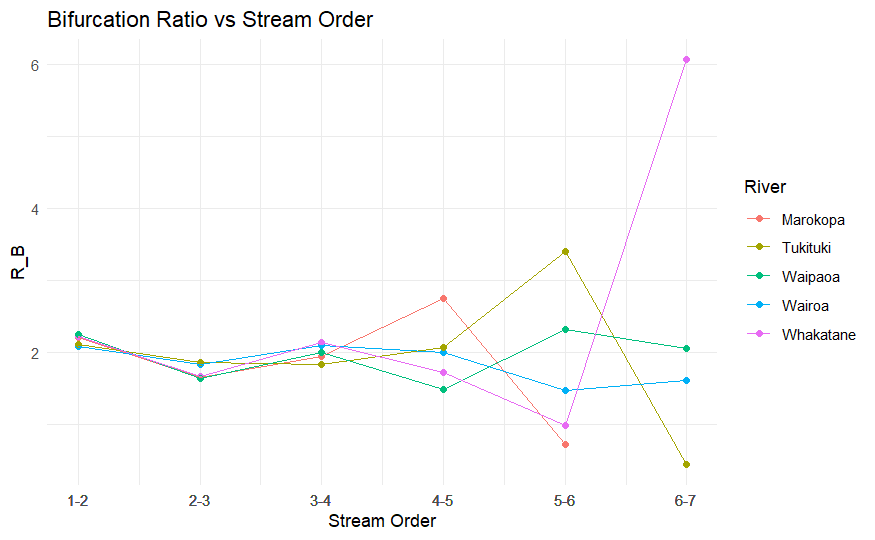
Our experiments to evaluate the impact of river connectivity and weight will be conducted with a fixed value of connectivity or weight. In this section, we compare the results of fixed connectivity and fixed weight.

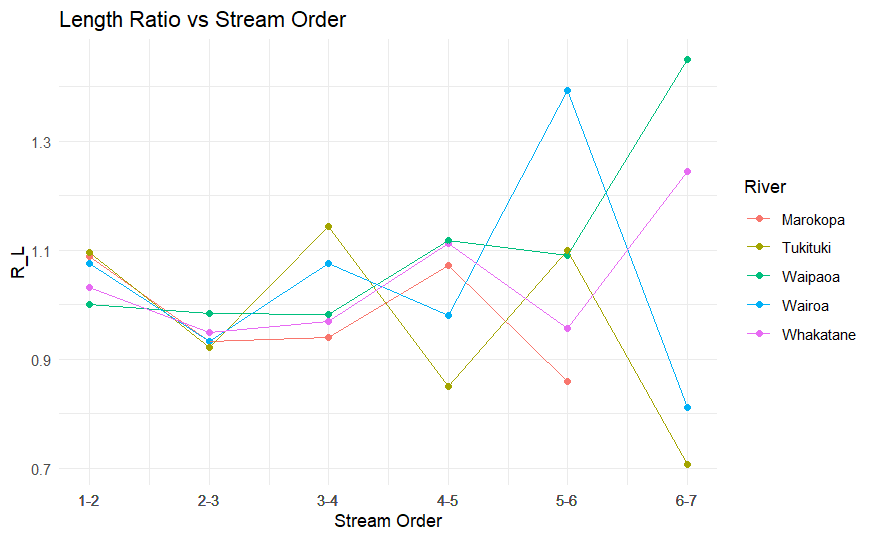
## 5.1.1 Connectivity Selection

重新画一下图，小的话一个图，最大的画一个图，最终ploty也展示

## 5.1.2 Weight Selection

## 5.2 River geometry





## 5.3 River topology

## 5.4 Summary

# Conclusions

# References

[1] Carraro, L., Altermatt, F. Optimal Channel Networks accurately model ecologically-relevant geomorphological features of branching river networks. Commun Earth Environ 3, 125 (2022).

[2] Hette-Tronquart N, Belliard J, Tales E, et al. Stable isotopes reveal food web modifications along the upstream-downstream gradient of a temperate stream. Aquatic Sciences, 2016, 78(2): 255–265

[3] Sklar L S, Dietrich W E, Foufoula-Georgiou E, et al. Do gravel bed river size distributions record channel network structure. Water Resources Research, 2006, 42(6): 770–775

[4] Shelef E, Hilley G E. Symmetry, randomness, and process in the structure of branched channel networks. Geophysical Research Letters, 2014, 41(10): 3485–3493

[5] Horton, R.E. 1945: Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology. Bulletin of the Geological Society of America 56, 275-370.

[6] Strahler A N. Quantitative analysis of watershed geomorphology [J]. Transactions of the American Geophysical Union, 1957, 38 :913-920.

[7] Tokunaga E. Consideration on the Composition of Drainage Networks and Their Evolution, Geographical Report [R].Tokyo, Japan: Tokyo Metropolitan University, 1978.

[8] Peckham S D. New results for self-similar trees with applications to river networks [J]. Water Resources Research, 1995,31 (4) : 1023-1029.

[9] LIU Huaixiang, WANG Zhaoyin. Morphological feature and distribution of typical river networks [J]. Journal of Hydraulic Engineering, 2007, 38 (11) : 1354-1357.(in Chinese)

[10] Rougier, N. P., Droettboom, M., & Bourne, P. E. (2014). Ten simple rules for better figures. PLOS Computational Biology, 10 (9), e1003833.

[11] Esri. "Esri Shapefile Technical Description." Esri, Inc. 1998.

[12] Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal, 10(1), 439-446.

[13] Benda, L., Andras, K., Miller, D., & Bigelow, P. (2004). Confluence effects in rivers: Interactions of basin scale, network geometry, and disturbance regimes. Water Resources Research, 40(5).

# Appendix A

# Some extra things

This is an optional chapter for any additional material that does not ﬁt conveniently into the body of the text (e.g., data, copies of computer programmes). Note that appendices won’t necessarily be read by the examiner.