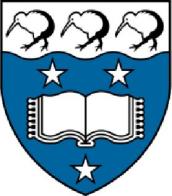
River Visualization



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].

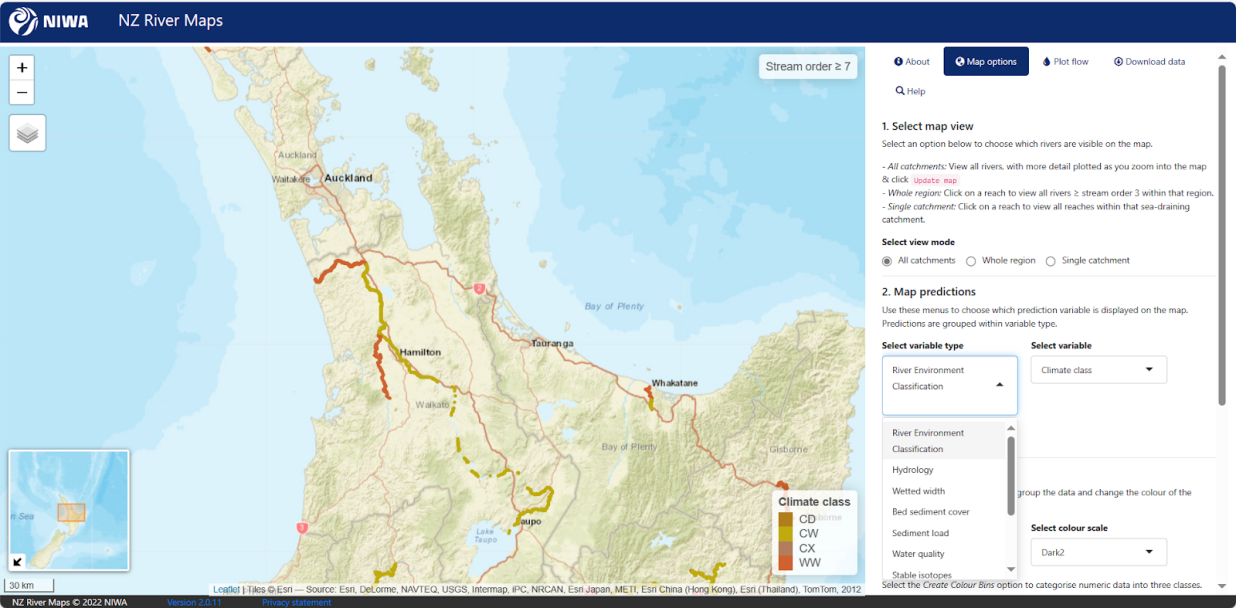


Figure 1. NZ River Maps of NIWA

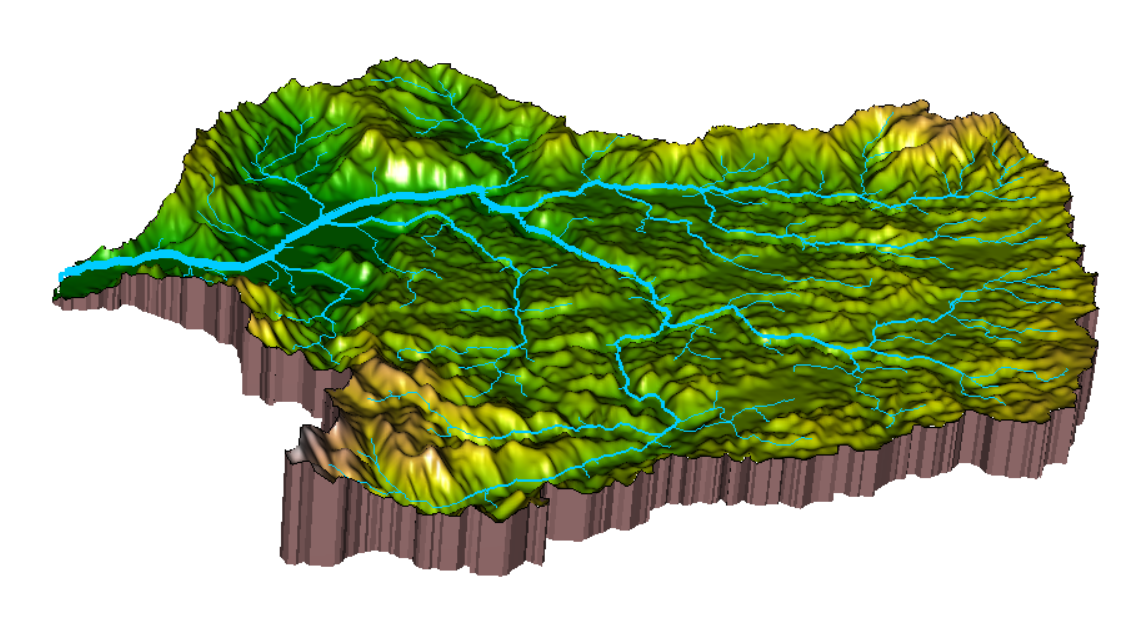


Figure 2. Visualisation of Optimal Channel Network Visualisation on Marokopa River

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 *i* can be statistically obtained, where *i* = 1,2,... , *N*. For Tianran 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.

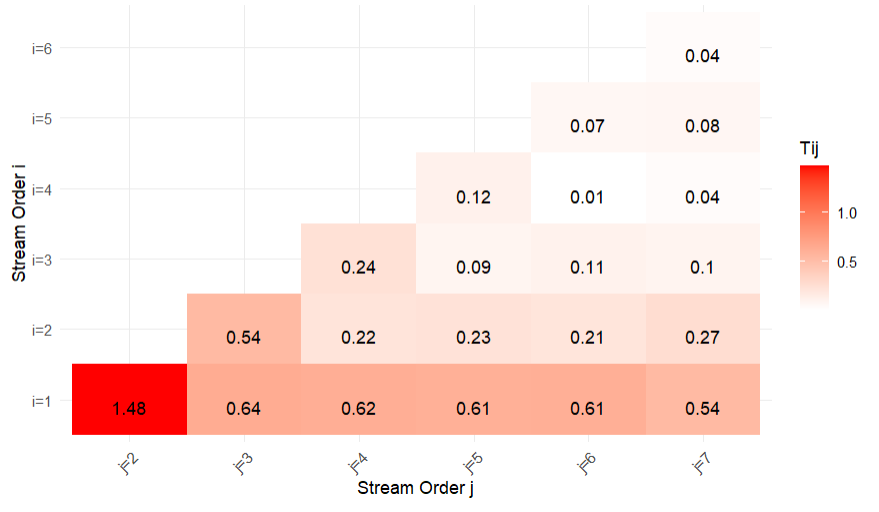


Figure 3. Side Branching Ratio for Waipaoa River

## 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.

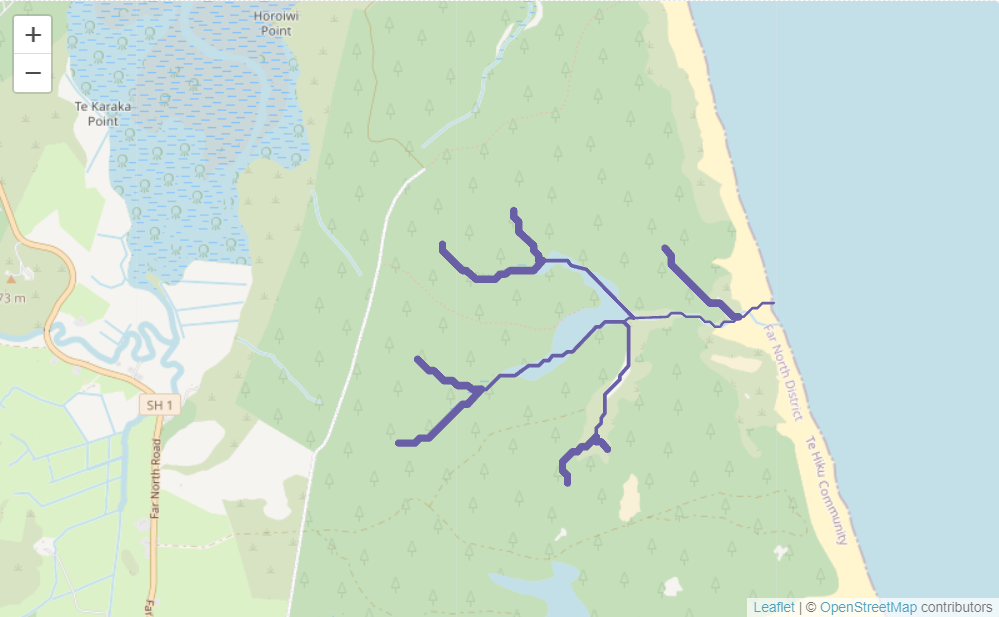


Figure 4 Example of Bad Visualization on Strahler Level of Kahiaka River

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.

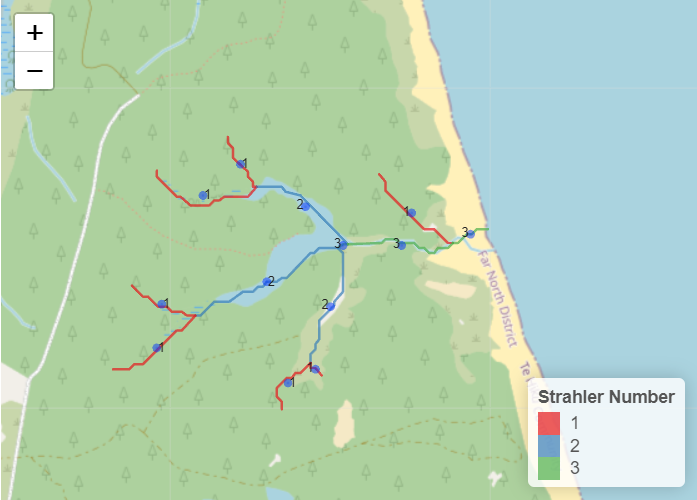


Figure 5. Example of Good Visualization on Strahler Level of Kahiaka River

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].

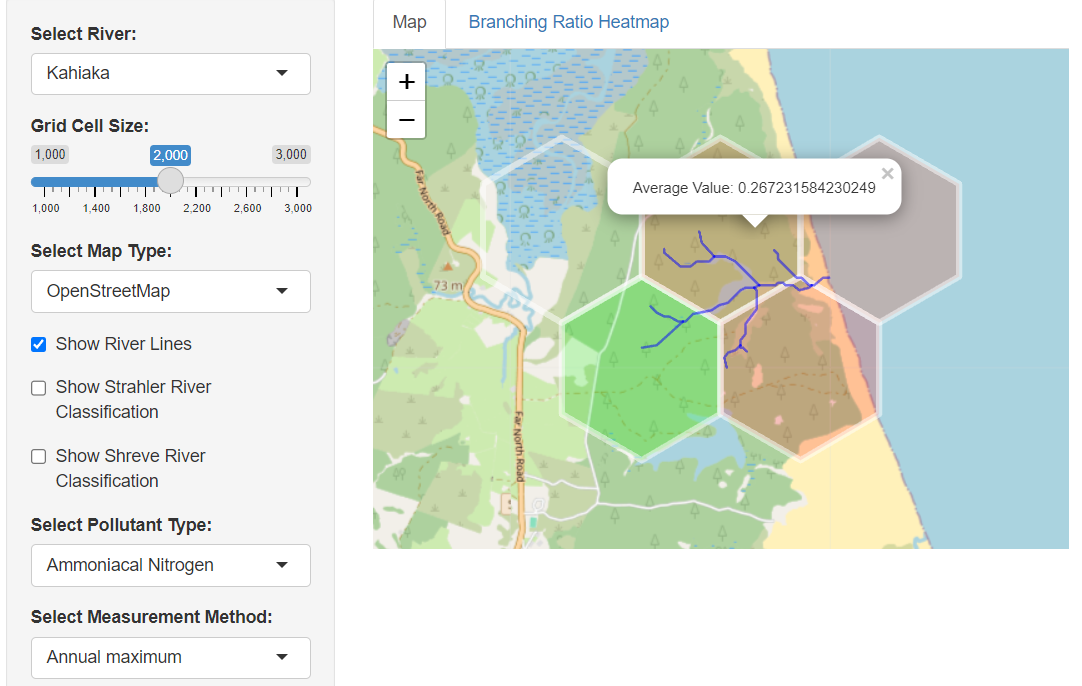


Figure 6 Example of River Visualization Project

## 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 will discuss different visualization techniques of how to implement visualization. We will also look at the Shiny package used to create the application.**这里没写完，等后面章节写了再补充**

# Chapter 2 Visualization Techniques and Shiny Application Development

This chapter discussed the techniques used for visualizing river lines and the Shiny tools used to create a interface to end-users. To be more specific, 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 main idea here is to summarize and analyze the advantages and disadvantages of several existing visualization methods, so as to learn from their strengths to complement their shortcomings and then use Shiny to achieve a unique way of data visualization. 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.

## 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.

## 2.1.1 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.

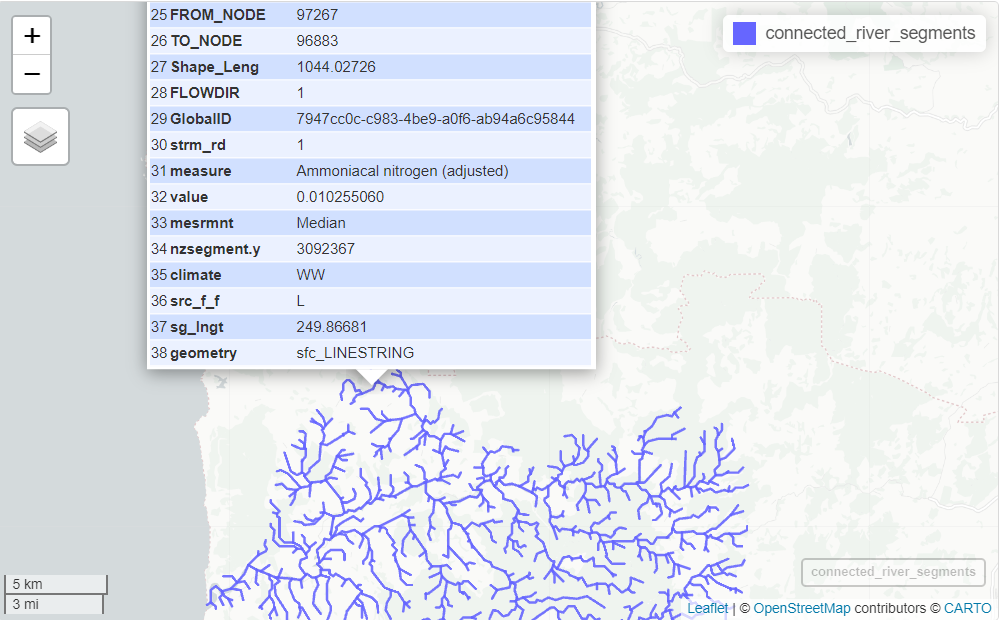


Figure 7 Example of `Mapview` interactivity

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

## 2.1.2 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.

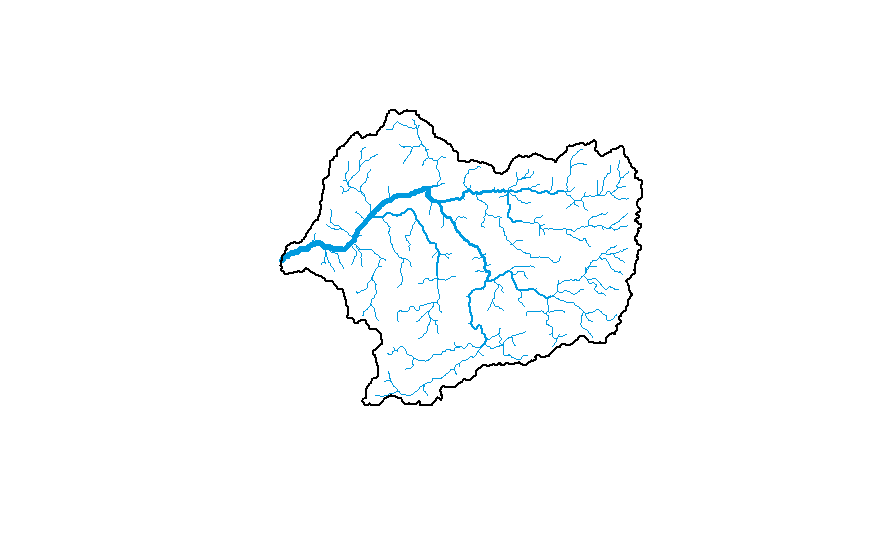


Figure 9 Example of Marokopa river Using OCN

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].

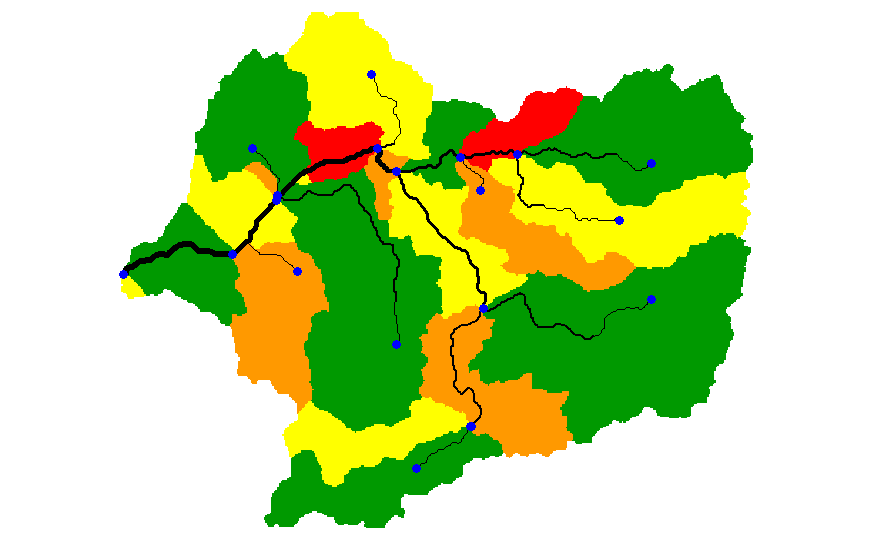


Figure 10 a. Subcathment of Marokopa river using OCN

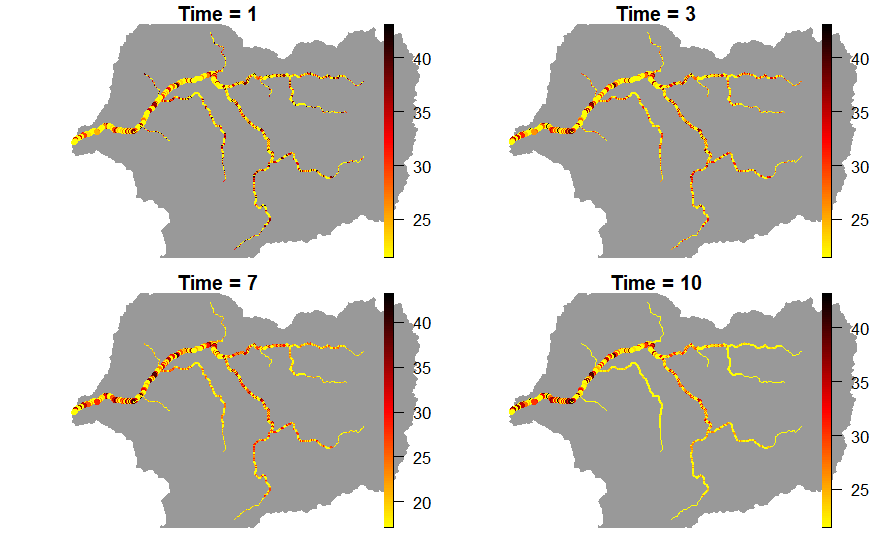


Figure 10 b. Metapopulation Simulation of Marokopa river using OCN

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)

2.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.

## 2.2 River Visualization 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.

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 |

## 2.2.1 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.

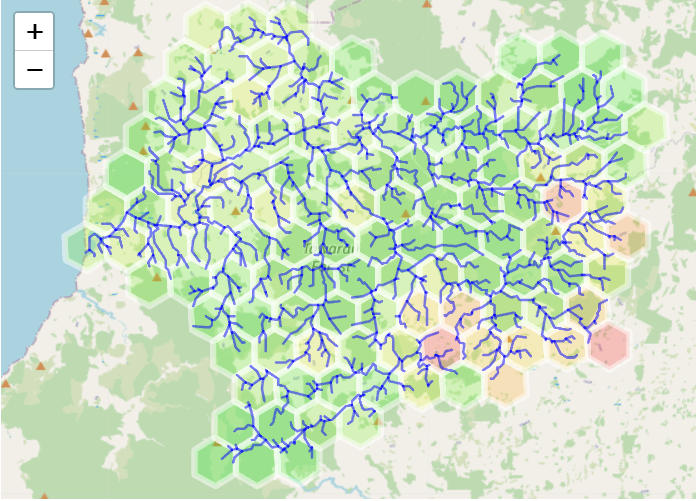


Figure 11a. River Visualization Shiny App River Lines Output

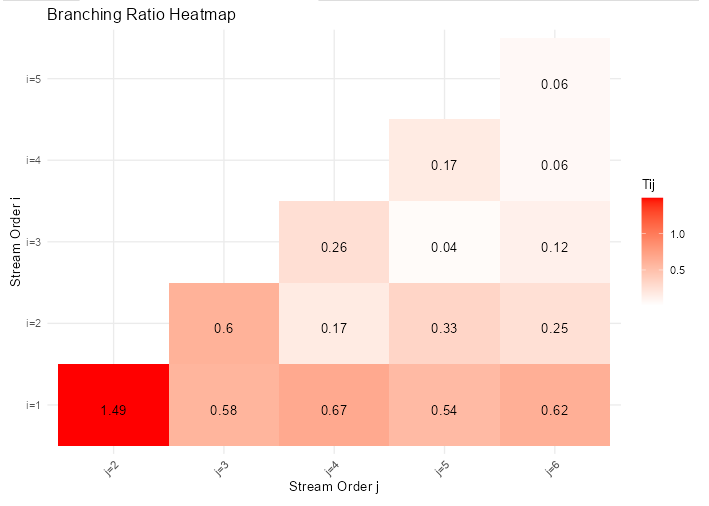


Figure 11b. River Visualization Shiny App Branching Ratio Output

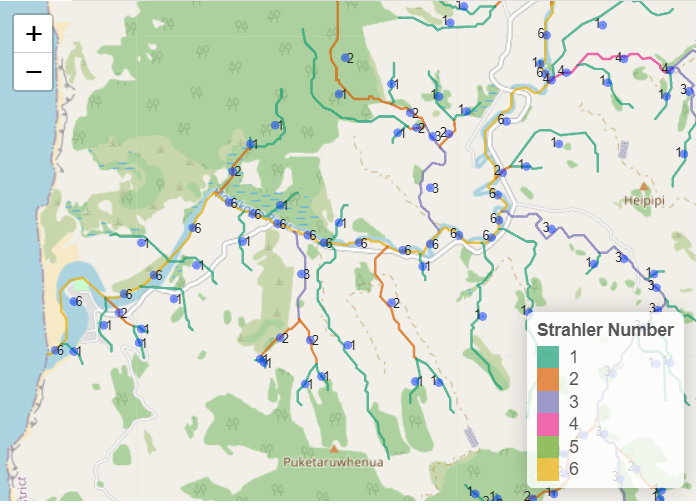


Figure 11c. River Visualization Shiny App Strahler Classification Output

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.

## 2.2.2 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) .

# Methodologies and analysis

2.1 Methodologies

2.2 Analysis

# Chapter 3

# Discussion

3.1 Main results

3.2 Discussion

# Chapter 4

# 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.

# 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.