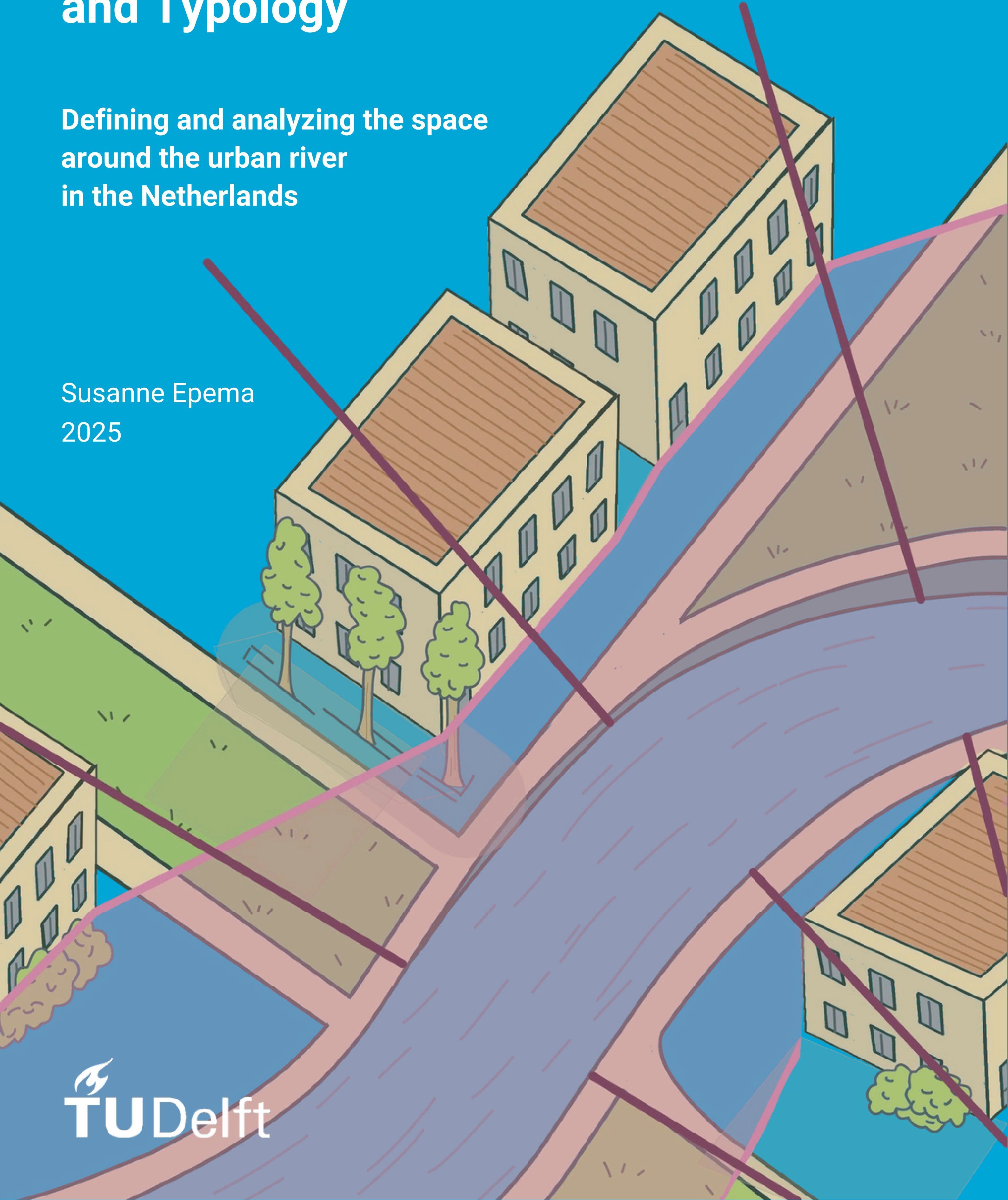
MSc Thesis in Geomatics

Semantic Urban River Space Delineation and Typology

Defining and analyzing the space around the urban river in the Netherlands

Susanne Epema

2025



MSc thesis in Geomatics

Semantic urban river space delineation and typology

Defining and analyzing the space around the urban river in the Netherlands

Susanne Epema

April 2025

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of Master of Science in Geomatics

Susanne Epema: *Semantic urban river space delineation and typology: Defining and analyzing the space around the urban river in the Netherlands* (2025)

The work in this thesis was carried out in the:

3D geoinformation group
Delft University of Technology

Supervisors:	Dr. Daniele Cannatella
	Dr. Claudiu Forgaci
Co-reader:	Dr. Giorgio Agugiaro

Abstract

The urban river space is the area in the city surrounding a river, distinguishing itself from other parts of the city by this relationship with the water. Urban river spaces around the world are increasingly under development and being regenerated. Urban planning solutions necessitate a comprehensive overview of this space, their boundaries and their characteristics. However, the concept of the urban river space is ambiguously defined, with varying definitions across studies. This thesis addresses the need for a standardized approach through semantic urban river space delineation to facilitate cross-case analysis. Three delineation methods are proposed and applied to urban areas in the Netherlands: the first building line based on visible building nodes, the visible space derived from viewshed analysis, and the floodable area, based on 100-year flood depth data. As urban river spaces are often represented as cross-sectional segments in research, this segment is used as unit to develop a typology of urban river spaces in the Netherlands. Properties of the segment, such as elevation, landuse, vegetation, flood risk, and visibility are quantified and used as input for the k-means clustering algorithm. 10 clusters are derived, each representing a semantic type of river space, resulting in a data-driven typology that enhances the understanding of urban river spaces in the Netherlands.

Acknowledgements

First of all, I want to express my gratitude to my supervisors Claudiu Forgaci and Daniele Cannatella, for their knowledge, advice, patience, and encouragement. Their insightful feedback, countless discussions on river spaces, and continuous support made this thesis possible. I would like to thank my co-reader Giorgio Agugiaro for his feedback which greatly improved aspects of my thesis, and Jantien Stoter for the feedback in the starting stage of my thesis, resulting in the development of the river space typology.

Secondly, I want to my parents and sisters for their support throughout my student life. A special thanks to Edith, for studying with me at the TU Delft and for helping me choose this master's program. I would also like to thank my fellow students during my master, especially Puti, for navigating every challenge of the Geomatics program with me. Going through the courses would have been a lot harder without your support.

A heartfelt thank you to Dylan, Floris, Niklas, Nathan and Prakhar for being there with me in the Geolab, studying (or dilly-dallying) with me, motivating me, and listening to my thoughts on river spaces, clustering algorithms and delineation methods. I want to thank all the other Delftsche Dodgers for keeping me active and grounded during my years at the TU Delft.
Finally, I would like to thank my boyfriend Per for his love, support and encouragement during the challenging times of my thesis.

Contents

List of Tables

List of Algorithms

List of Figures

Acronyms

1. Introduction

This chapter will introduce the topic of urban river spaces. The motivation for this thesis is given, specifying the usability of the delineation and a data-driven typology. Some specific use cases are examined to contextualize the problem. Afterwards, the research objectives and scope of this thesis are specified.

1.1. Background and motivation

Rivers play a pivotal role in urban environments, shaping cities' spatial, ecological and cultural identities. Historically, most urbanisation begins at locations that provide in services and key resources such as food, fresh water, ease of defense, waste disposal and transportation links [Grimm et al., 2008]. In general, rivers are able to provide all of these, and therefore many cities around the world are build around them. Rivers play a key role in shaping the city identity. They play a crucial role in city planning, the availability of recreational spaces, flood control and ecological benefits within the city [Prominski et al., 2023]. Rivers connect urban spaces, contribute to space identity, and reflect city characteristics [Pattacini, 2021].

(a) Rivers (BGT dataset and orthophoto via Landelijke Voorziening Beeldmateriaal)

(b) Buildings and canals (Image source: Commons [2025])

Figure 1.1.: Amsterdam's canals

Pre-industrial cities had a complex relationship with water, viewing it as an integral part of the urban landscape rather than a limitation. This connection was disrupted with the advent of the industrial era [Moretti, 2008], in which access to the water was blocked by port activities, and a decrease in importance of waterborne transport as cities expanded into mainland and the development of railroads. Urban rivers are distinct from their natural counterparts, due to their direct interaction with the cityscape and the challenges, but also opportunities, that this relationship entails. Contemporary urban challenges, such as climate change, rapid urbanization and competing interests in land use have intensified the pressures on river spaces. The intensity of development, characterized by the competing interests, often leads to standardization, areas without established identity, fragmentation, loss of heritage, loss of functional and ecological diversity, the loss of ecological areas, and an increased sensitivity to flooding [Petrtýlová and Matej, 2022]. These dynamics and resulting complications necessitate innovative approaches to planning, managing and analyzing urban river spaces. Urban rivers differ tremendously from natural rivers due to their direct connection to the urban environment, which forces the river to adapt to the needs of the city. Therefore, urban rivers and their spaces have an artificial component to them, in addition to their natural dynamics. For example, Amsterdam (figure 1.1) is known for its canals, which shape the city's identity, but are highly artificial to serve the city's needs. Natural river landscapes consist of dynamic mosaics of habitat patches distributed along environmental gradients, shaped and interconnected by disturbances from flow and sediment transport, often reinforced and enhanced by vegetation growth. Urban development introduces significant changes to the form and function of river systems. It alters all the fundamental processes that shape river corridor dynamics and bio-complexity [Gurnell et al., 2007]. Richardson and Soloviev [2021] introduce the concept of *Urban River Syndrome*, by which they mean the accelerating deterioration of urban river ecology as a result of urbanization. As cities expand, natural waterways are altered in ways that negatively impact their ecosystems and hydrological functions.

Recent trends in urban river management highlight the growing recognition of their ecological and cultural significance, as attention shifts to cope with climate change: the importance of resilience of the city and enhancement of natural infrastructure in urban areas [Brown, 2020], as well as the growing popularity of recreational activities [Moretti, 2008]. The potential of the natural dynamics of the river is being appreciated, and urban planning of the river space moves from the completely controlled, artificial river to a river with its own dynamics. An increased attention is given to the benefits the natural river brings. Water is once again seen as a resource [Moretti, 2008]. Increasing attention is given to water in the city from an urban planning point of view [Prominski et al., 2023], redeveloping the urban riparian landscape. Port regeneration, city beaches, waterside living, and new riverside promenades are being developed , as ways of improving quality of life.

An example of the reversed trend going from controlled to a river with its own dynamics

(a) Before restoration. Canalized, artificial river.

(b) After restoration. Re-natured, lowering the flood risk. More natural river flow.

Figure 1.2.: The Isar in Munich, Germany. (Image source: City of Munich [2016])

is the river Isar in Munich via the *Isar plan*, in which an 8 kilometer stretch of the river in the city has been re-natured (see figure 1.2). This river has undergone significant restoration efforts as part of a comprehensive flood risk management plan using a Nature Based

solution (NBS) approach. This approach has proven effective in mitigating flood risks while improving the river's ecological status [Pugliese et al., 2022]. The project, implemented in the early 2000s, addressed social, economic and ecological factors influencing urban river park design [Zingraff-Hamed et al., 2022]. The water quality of the river had been improved to the point where the water meets bathing standards and thus is nowadays used as a bathing site in summer.
The potential that can be reached in development or regeneration of the urban waterfront is unrealized by underestimation of the river-city relationship and place identity, and uncoordinated development of the waterfront as a whole [Petrtýlová and Matej, 2022]. Designers of river spaces must balance flood control, open space design, and ecological requirements [Prominski et al., 2023] in their attempts to regenerate or develop these spaces.

But what exactly is the river space? What determines the boundary of this unique area in the city? No structured delineation of the space exists, even though this is of great importance in terms of planning and analysis to assess the current state. If the goal is to compare two river spaces, we want to make sure we compare likes with likes to make this a fair comparison. Additionally, to be able to understand anything about fundamental urban questions, appropriate and meaningful definitions are required, and the delineation of space is fundamental for policies targeting specific geographies [Duranton, 2021]. Delineation involves providing a detailed representation or precise boundary of a particular space, emphasizing clarity and specificity. It serves as a foundational tool for understanding and characterizing spaces, making it particularly relevant for urban studies. The significance of delineation of urban boundaries is well-documented in research. For instance, Dosary et al. [2002] highlights its role in urban management, while Burian et al. [2012] underscores the necessity of objectively defining urban centers to enable consistent statistical comparisons across cities. However, these studies also point out the challenge of developing boundary-determining methods that remain applicable beyond the specific cases they were designed for. Thus, the creation of objective and transferable frameworks for boundary determination is essential to facilitate meaningful comparisons between cities.

Section 2.1 will discuss the observation that river spaces lack a universally agreed-upon definition, and their characteristics often vary based on the urban context. A city can be very complex. There is a need to define areas and make distinctions between them in order to analyze and manage them effectively, and delineation can provide a base structure for comparison and analysis, enabling the identification of spatial patterns and relationships. This also shows a challenge in choosing a delineation method. How can it be made applicable to any city to make it suitable for cross-case analysis? What is the boundary based on? This is also where semantic information plays a role in the delineation. The delineation of the river space needs to be tied to interpretable semantic information to be able to make sense of what is there.

The integration of elevation and 3D spatial data into delineation methods can enhance the depth and accuracy of spatial analysis. This shift from 2D to 3D representation reflects broader advancements in geo-spatial technologies, driven by the increasing availability of detailed 3D datasets, such as a Digital Surface Model (DSM) and Digital Terrain Model (DTM). Incorporating 3D data introduces new opportunities to address urban complexities, for example by improving the spatial understanding through identification of vertical obstacles and changes in elevation. The enhance spatial understanding aids more informed decision making in urban planning applications. Connectivity characteristics such as water accessibility and visual permeability of the urban river can only be determined using 3D data.
The expanding availability of open accessible data and new tools and technologies emerging

constantly offer the opportunity to test and create new analytical methods and processes. This allows researchers to offer evidence-based knowledge to guide action [Fleischmann et al., 2022]. The river space is often described and classified in an observational manner, such as by Prominski et al. [2023] and Durán Vian et al. [2021]. These methods do not take full advantage of the (big) data and emerging technologies available. There is very little data-driven insight applied developing these river space typologies. Here lays an opportunity to explore the process of using quantitative open spatial data to compare river spaces and gain a comprehensive understanding of the space, which may reveal 'hidden' properties which would be difficult to spot in an observational manner.

To summarize, the urban river space is under construction as a result of neglect and human interference in the past. The lack of structured, detailed and comparable units of analysis for spatial decision-making creates significant gaps in effective planning. Analyzing reliable quantitative data can provide a stronger foundation for comparing and analyzing river spaces in urban context.

1.2. Employing river space delineation and typology

As mentioned before, the delineation of space is important in order to be able to make fair comparison between spaces, to allow us to understand anything about fundamental urban questions, and is fundamental for policies targeting specific geographical areas. A structured, semantically enriched delineation opens doors for cross-case analysis between cities, to determine qualities of the space, and in terms of urban river space regeneration, offers the ability to define the space for targeting policies.

A typology of river spaces provides an understanding of the space in its relation to other river spaces. Such a typology has practical applications in two main ways:

1. **Descriptive applications**: By categorizing river spaces based on their current characteristics, one can assess their roles and behaviors within the urban context. Identifying types of river spaces can inform urban analysis and planning. For instance, Durán Vian et al. [2021] in their classification of urban riverfront walks and parks, state that organizing riverfront recreational areas can be a useful tool for geographers, architects, planners, engineers and other professionals while studying, planning, and managing riverside open spaces. Prominski et al. [2023] describes existing types of river spaces to serve as a design tool for urban planning.

2. **Normative applications**: Typologies can also guide decision-making by defining the desired or potential states of river spaces. They can help identify the interventions needed to transform a space into its ideal state, considering its ecological, social, and urban roles.

1.2.1. Urban planning use cases

This section discusses some cases in which a data-driven typology of river space and a standardized delineation of the space add value in terms of adding a data-science perspective on observational studies, or where a semantic delineation can structurally determine the space analyzed.

Assessment of place identity

Stating that the development of waterfronts is often uncoordinated which is causing them to not reach their full potential, Petrtýlová and Matej [2022] attempts to formally establish the unique place identity of the river and the river-city relationship by defining character areas along the Bratislava waterfront. A vision for each space was formulated based on the defined character elements. This definition was a result of a combination of social-spatial analysis, a comparative analysis, and an online questionnaire survey. The comparative analysis consisted of a rating system for ecosystem quality, program content, identity, maintenance and safety, and ease of movement. Elevation or 3D data is not used to determine qualities of the space. Character areas here are seen as long stretches of land neighboring the river, and the boundaries of these segments are vague and defined based on the specific case of the Bratislava waterfront. This makes this study and the used definitions very case-specific, which prevent comparison with other rivers. The character areas are only defined in terms of the this particular city. Studies like these can therefore benefit from delineation of the river space, to enhance the ability to scale up for cross-case analysis, to ensure a valid comparison between cities. This is an example of a normative application of typology, as areas of different type are defined and analyzed to create a urban planning vision for each type to improve the space. A structured, data driven typology, with possible added variables, based on a large number of river spaces can aid in assessing the current state of the space and possibly determining shortcoming if the current state is not of desired type. Then the desired type can be selected and an urban planning strategy can be determined.

Temporal assessment of regeneration projects

Sopena Porta and Pellicer [2024] performed a comparative study on two urban planning and landscape projects on urban riverbanks of the Ebro river in Zaragoza, Spain, and the Isar river in Munich, Germany. Here they select a number of actions taken in the two projects, which remodeled appearance and improved the conditions. This is an example where river space typology can aid the comparison and assessment of improvement. Similarly as done in this study for initial- and post-scores for quality levels, types can be determined before and after regeneration efforts to assess changes and the state of the space in comparison to the general characteristics of river spaces.

The river space as recreational area

Durán Vian et al. [2021] offers a classification system to describe open space classes. From a literature review and analysis of cases a classification framework was derived. This was done through fieldwork, where each city was visited. However, many of the indicators in this classification could have been derived from data instead of site visits. Data exists for indicators such as the shape of the cross-section following elevation data, vegetation or pedestrian paths. Therefore, such a classification can be automatized or scaled up by a more data driven approach. Similarly, Rahman and Kautsary [2024] primary uses a field observation approach for their river parks typology. Their secondary observations were done through big data analysis through google street view, but only to locate the parks and to obtain a view of the space. No elevation data or automatic extraction of indicators was applied.

1.3. Research objectives

The overarching goal of this research is to delineate urban river spaces, to enrich the description of the space with semantic information, and to move towards a scalable typology of the urban river space. This typology then serves as a robust framework for understanding and managing these spaces.

Therefore, the primary objectives of this research become

1. Delineation of the river space by refining existing methods with 3D information, enhancing detail and utility.
2. The semantic enrichment of the description of the urban river space, adding meaning and structure to spatial data for better decision-making.
3. Typology development in order to establish a system to classify and analyze urban river spaces.

To achieve these objectives, this research calls to answer several key questions

How can urban river spaces be delineated and analyzed using (2D and) 3D spatial data?

- What adaptations to traditional 2D methods are necessary to add value?
- What areas within the river space can be distinguished?

What types of spatial configurations and qualities can be observed in urban river spaces as seen through 3D delineation and spatial analysis?

- How do river spaces differ spatially?
- How can properties be derived from the role of the river space in the urban environment and how can they be quantified?
- What clusters of co-occurring properties can be found when looking at multiple river spaces?

1.4. Research scope

This research focuses exclusively on river spaces within urban environments, with a particular emphasis on the Netherlands as a proof-of-concept case. This implies that types of datasets are used that may not be available in other countries.
The scope deliberately excludes certain aspects to maintain clarity and feasibility. For example, the study does not consider detailed river morphology or environmental parameters like pollution, focusing on the additional value that lays in spatial geo-data, and thus focuses on spatial configurations. This selective approach allows for more targeted analysis, but suggests areas for future research broadening the scope by integrating additional information to the description of the river space.
No assessment of the river space is done. The goal is to investigate the space, find ways to define it, represent them and to analyze aspects of it in order to develop river space types. Nothing is said about the quality of the types.

Only the current situation of the space is taken into account. No temporal data is used in the sense that it is not compared to former situations.

2. Related work

This chapter discusses relevant research and theory regarding urban river space delineation and typology. Section 2.1 discusses existing river space definitions and the lack of universal language when it comes to analyzing this space. The only existing river space delineation method is discussed, and ways to enhance this method. Existing river space typologies are described, and the cross-sectional visualization of the space in literature. Finally, key properties of the river space are discussed in section 2.3.

2.1. Defining the urban river space

The boundaries of the river space, waterfront, river front, water edge, city port, and similar terms, and the differences between these, is not universally defined in literature. Meaning differs between authors, research objectives and definitions have evolved over time. Petrtýlová and Matej [2022] states that the borders of the waterfront area are usually not clearly defined, in their research on waterfront character areas to establish the city-river relationship and a unique place identity. Here the waterfront has no clear spatial boundary, but is perceived as the space in close contact with the river combined with the adjacent area. Excluded is the remaining space due to it being occupied, obstructed or separated by flood barriers, buildings and elements of technical and transport infrastructure. Bisjak [2005] defines the river front as ”the area between the river and the first line of buildings, including these buildings”. Moretti [2008] states that ”the word ‘waterfront’ identifies the urban area in direct contact with water. In cities on water, this area usually corresponds to the area occupied by port infrastructures and port activities.” Durán Vian et al. [2021] states that meaning of the ’waterfront’ differs between authors and eventually defines the riverfront as ”the transition area between the river and the city rather than a direct contact zone”.

Definitions differ on objectives of research and what authors need from the space, and a clear spatial delineation is often not given. This lack of universal language within the different disciplines causes the results of research to be disconnected. Forgaci [2018] developed a spatial delineation method for the Urban river corridor (URC) and parts within it, including the river space. Here, the river space is delineated by the first line of buildings. Boundaries of the waterfront or areas within the river space are not defined. This delineation neglects spatial elements, land cover differences, green spaces and obstructions within the river space are neglected in this delineation method, while these features are of importance to gain a full understanding of both geometry and qualities of the space. A structured semantic delineation method of the space based on spatial characteristics of the area is needed to add detail and make research transferable across disciplines.

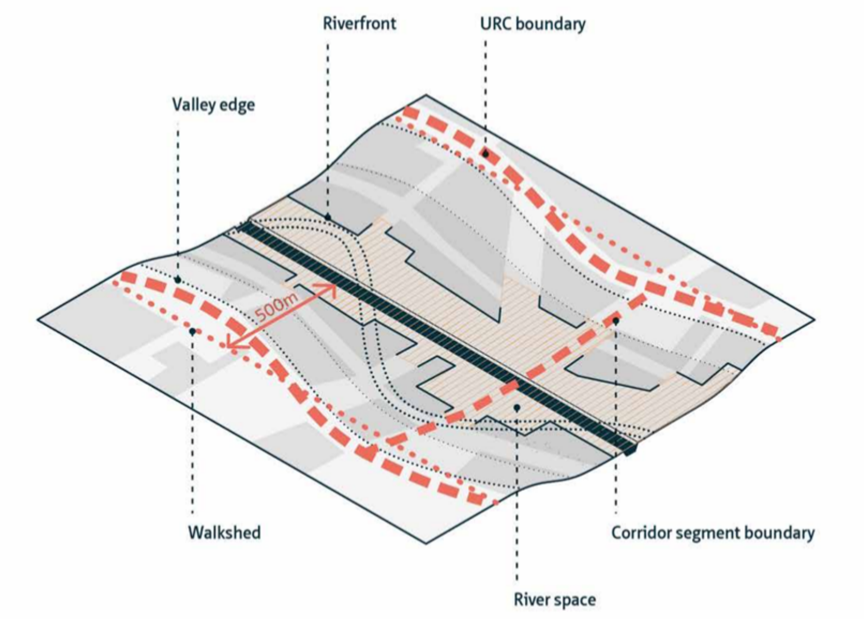


Figure 2.1.: The 2D spatial delineation method of the URC [Forgaci, 2018]

2.1.1. Delineation of the corridor

The delineation method for the URC by Forgaci [2018] is a 2D method It draws the boundary based on presence of buildings, but does not use any 3D details of the area. The definition here is given from a spatial-morphological perspective. It entails the spatial structures that help integrate the river within the urban environment. This implies it takes into account certain components of the relationship between the river and the city, such that any property of the URC concerns both the river and the city. The delineation that considers such configuration and composition of the urban fabric, has an underlying social-environmental logic that is better aligned with the boundaries of a naturally occurring phenomenon than an arbitrarily chosen area. Figure 2.1 shows an overview of the 2D delineation. The components are defined as following:

- **Valley edge**: Determined from, for example, a Digital Elevation Model (DEM) using a method of river corridor delineation.
- **Walkshed**: An extension of the outer boundary, in this definition towards 500 meters.
- **Riverfront**: The space along the river delineated by the built front.
- **URC boundary**: Determined by the main roads parallel, next to and outside the river valley.
- **River space**: The direct contact area between the river and the first line of buildings, including these buildings. Beige coloured in the image.
- **Corridor segment boundary**: Divide the URC along major transversal traffic lines, most of the times when there is a bridge.

- **River**: Thick black line is the canal and the curved line is the natural trajectory of the river.
- **Streets** are shown in light gray, **buildings** in darker gray.

This delineation method defines the two-dimensional river space boundary as "the direct contact area between the river and the first line of buildings, including these buildings" taken from Silva et al. [2004]. Although this definition seems clear enough, the method is underdeveloped in the sense that there are undefined characteristics of this boundary. What is considered the first line of buildings? If there is a gap between two buildings, does the delineation just connect the two or does it search for a further building as connection? How far away could this building be? Are obstacles and roads considered bounding the direct contact area? I works if manually extracted such that an assessment by the researcher can be done in these cases, but not when automatically extracting this boundary. No structured methodology is given to adapt to more complex cases.

2.1.2. 3D enhanced river space delineation

The approach discussed in the previous section overlooks critical spatial characteristics within the space. Vertical changes—such as flood walls, trees, or other barriers—can significantly impact the continuity and connectivity of the river space but are not accounted for. Buildings behind smaller structures or levees, may have important implications for spatial accessibility and functionality yet remain unrepresented. Other characteristics, such as terrain elevation changes and floodability of the space are not taken into account. Therefore, methods to include this extra information have to be investigated.

The work *River Space Design* [Prominski et al., 2023] offers a comprehensive overview of design tools for the river space. Two boundary lines are given within each design tool: the flood limit line, which is especially of importance to determine if a specific object or structure falls within the flood limit or outside of it and is thus vulnerable to floods or not, and the boundary line for the river's self-dynamic, specifying the maximum reach the river is allowed to have by the structure of the design (figure 2.2). These boundary lines are thus influenced by flood control and the natural dynamics of the river itself. The addition of these two boundaries offers extra detail to the space, aiding a greater understanding of its characteristics. A characteristic of the river space is its flood-ability. Flood barriers may influence

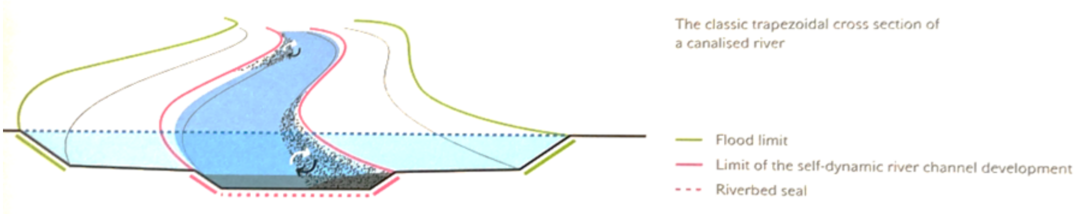


Figure 2.2.: The flood limit and limit of the self-dynamic of the river channel development by Prominski et al. [2023]

accessibility to and visual connectivity with the river. Floodplains can lay within the urban river space, and thus will periodically be filled with water. Durán Vian et al. [2021] uses the floodplain to differentiate between urban river spaces and valleys. Silva et al. [2004] defines the flood vulnerability of the corridor as the percentage located within the area of 100-year flood event and the number of people living in the floodplain. Delineating the flood line can therefore aid identifying particularly vulnerable areas around the urban river.

Forgaci [2018] discusses the visual permeability of the space as a indicator for spatial capacity. This is considered as the percentage of visible open space within the river space. Prominski et al. [2023] mentions the sight lines, visibility of the space and the view on the river in their typology, and has the concept of open space as fundamental. Apriliani and Dewi [2020] has physical and visual access to the riverbank, along the riverbank, and to the water considered in their study on the Cisadane Riverside as a parameter of accessibility. Visual permeability is considered an indicator for urban space quality by Bisjak [2005] in terms of aesthetic value. Durán Vian et al. [2021] has visual barriers between the water and open space as a factor of water accessibility. Visual access can be differentiated from physical access to water [Silva et al. [2004], Kondolf and Pinto [2017]]. Therefore, visual connectivity is deemed important for a range of reasons, including in terms of aesthetics, open space quality, spatial capacity, water accessibility and social connectivity. Here lays an opportunity for delineation, as the boundary of visible space around the river. This ensure a space with direct connection to the river

2.2. River space typology

A typology involves a combination of attributes that collectively define a type Djokić [2010]. Selecting criteria for defining a type requires careful consideration. Relevant parameters must be prioritized, and irrelevant ones excluded, to avoid distorting the structure of the classification. The usability of a typology depends on balancing the number of criteria and their complexity, as well as recognizing that the significance of each criterion may vary.

An unlimited amount of parameters can be defined to describe the urban river space. For example, spatial characteristics such as width of space and urban integration, physical aspects like the structure of riverbank, but also soil composition or the effects of the river dynamics, and ecological indicators related to vegetation and green patches. These indicators can be chosen based on the use of the typology. A typology based on aesthetic value and a typology based on flood vulnerability require different sets of indicators. However, some characteristics would be applicable to almost every application. For example, the elevation differences within the space are fundamental to a range of indicators, and therefore is always a significant criterion. As this thesis does not develop a typology for any specific application, the aim is to identify such indicators and use those to build the typology on. This would result in a *base* typology of river spaces in the Netherlands, and can be expanded depending on specific applications.

2.2.1. River space typologies in literature

Research regarding river space typology and ways to define such types exists in literature, usually for very specific use cases, but not always. Prominski et al. [2023] acknowledges the importance of differentiating between river spaces in terms of urban design. The work *River. Space. Design.* provides a typology of the urban river space in terms of three thematic fields: ecology, flood prediction and amenity. The typology is divided into different process spaces and shows an extensive list of design tools.
A typology based on a more specific problem is the influence of the land use and coverage on river water quality. In a study on the Pepe river in Surakarta City by Rini et al. [2020], spatial analysis was used to develop a typology and describe riverbank characteristics. The typology is based on the potential of the surface water to be discharged into clean water. The typology is thus based on infrastructure, land use and building coverage on the riverbanks. The study here extended to a radius of 700 meters around the river, as this buffer area was considered to influence water quality. Five river types are distinguished and used to formulate interventions for better river water quality.
As more spatial data becomes available for urban environments, urban form studies are able to utilize a data-driven approach for classification. Several studies have been conducted using clustering algorithms to represent urban form typologies. For example, Berghauser Pont et al. [2019] uses statistical clustering methods to present typologies for three key elements of urban form (streets, plots and buildings). Gil et al. [2012] presents a method to support the description and prescription of urban form using k-means statistical clustering technique to produce objective classifications. Wang et al. [2023] uses clustering to obtain urban types, and argues for the mapping of interpretable urban morphology of basic urban elements as an intermediate step instead of mapping non-physical aspects directly in research about, for example socio-economics. A clustering algorithm is applied to meaningful metric to obtain urban types which can be used to interpret other patterns, including socio-economic ones.

2.2.2. Cross-section segment representation

The choice of segment is a challenge that arises when developing a typology for the river space. The space along the river is a system that should be considered in its entirety. However, to develop a sensible typology, a unit of analysis has to be defined. In the lateral dimension, this can be achieved using the delineation methods or by using a pre-defined length of the segment. However, longitudinal segmentation is more difficult. What defines the boundary between types? If the segment chosen is too large, it most likely has different subspaces with different functions and spatial characteristics in it, and therefore could include multiple types within itself. This could be a problem when, for instance, using bridges as longitudinal bounds along the river. A solution to this problem is using cross-sectional

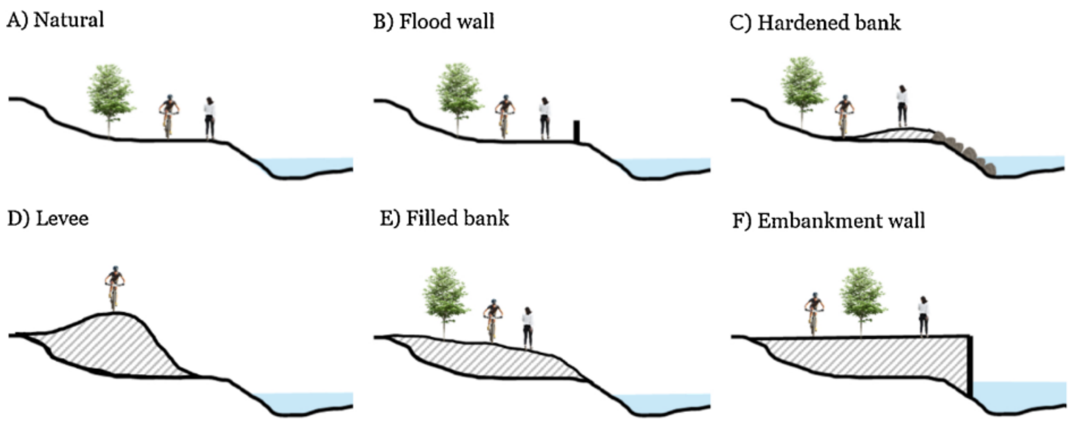


Figure 2.3.: Cross-section visualization for classifying urban riverfront parks and walks (Image source: Durán Vian et al. [2021])

segments. This is often done in literature analyzing properties of the river space, mainly as a way to visualize the state of the space or to describe a type of space. The river space at a certain location can largely be defined by the cross-sectional elevation values combined with semantic information on land use and other properties, and the continuity or dis-continuity between the cross-sections. Durán Vian et al. [2021], in their article for classifying urban riverfront parks, depicts the valley shape and common water edge barriers as cross-sections, including the elevation but also where it is altered by human, human accessibility and potential trees (figure 2.3). The elevation profiles of the section characterize the different water edge barriers. The profile together with the building locations show the urbanized valley shapes. In a study on the Cisadana riverside that required analyzing elements that constructed the riverbank as public facilities toward sustainable pillars by Apriliani and Dewi [2020], the situation is depicted in different zones of the river space as cross sections illustrating the elevation differences and obstacles in the terrain, as well as vegetation and soil composition (figure 2.4).

The typology for river spaces developed by Prominski et al. [2023] illustrated every process space, but not only as a cross-section. The design tools that fall under each process were often illustrated as a cross-sectional drawing, such as A1 in figure 2.5. Figure 2.6 however shows a2.1 and a2.2 longitudinal from the river. If we would take the cross-section lateral

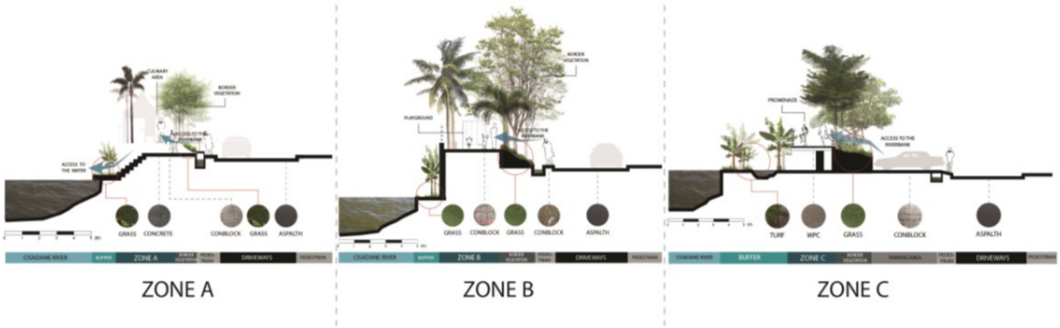


Figure 2.4.: Conditions in defined zones visualized as cross-sections (Image source: [Apriliani and Dewi, 2020])

from the river as done with the other design tool section A1, similar spatial characteristics would be seen. The difference here lays in the continuity of these characteristics along the river of the design tools.

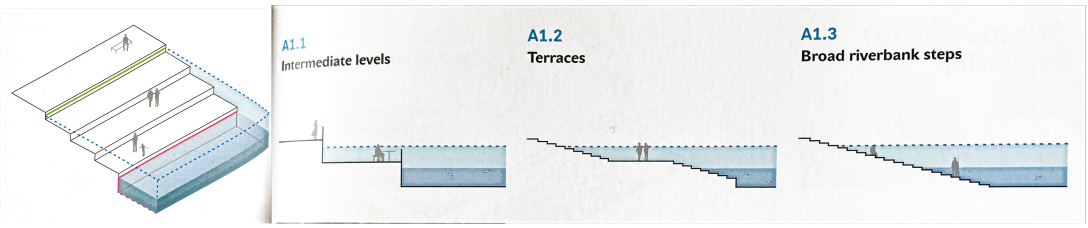


Figure 2.5.: River space typology types are visualized as cross-sections for embankment walls and promenades. (Image source: Prominski et al. [2023])

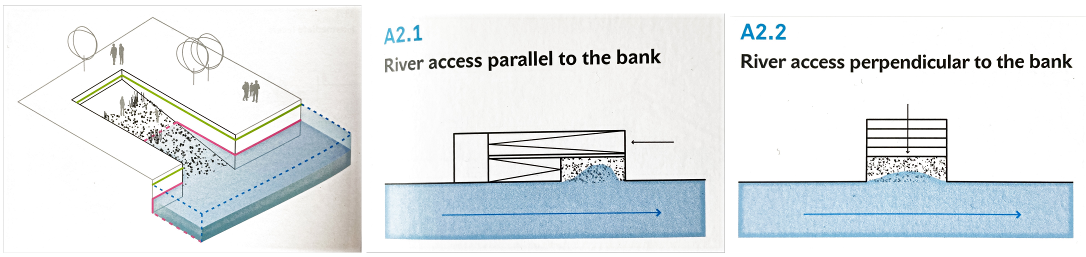


Figure 2.6.: River space typology types are visualized as cross-sections for embankment walls and promenades. (Image source: Prominski et al. [2023])

2.3. Characteristics of the river space

	Longitudinal	Lateral	Vertical
Ecological	Migration of species and flows of materials up and down the stream	Interactions with the watershed, geomorphology, and material and species movement between water and land	Exchanges between river and groundwater
Hydrological	Headwater-estuarine flows	Riparian-floodplain interaction	Riverine-groundwater relation
Social	Activities that run along the river, such as navigation or riverside traffic corridors, ranging from fast to slow movement	Visual and mobility connections (accessibility) (1) across the river and (2) transversally to and from the surrounding urban fabric	The direct interaction between people and water such as swimming, walking along embankments and dynamic floodable areas
Spatial	Continuous access along riverbanks for both people and ecosystem agents	Transposability of the river and accessibility from the surrounding fabric	Channel section configuration to allow access to and from water

Table 2.1.: Integrated three-dimensional connectivity. Table retrieved from Forgaci [2018], citing May [2006], Tetzlaff et al. (2007), Kondolf and Pinto [2017] and Gordon(1996)

Forgaci [2018] performed a trans disciplinary literature review on urban rivers, in which key principles were identified under four domain families: the environmental-ecological dimension, the social-economic dimension, the planning-governance dimension, and the spatial-morphological dimension.
The principle connectivity is the property that describes processes, movement and interactions within and between spaces in the URC, and is synthesized in three dimensions: longitudinal, lateral, and vertical based on the review of connectivity in urban rivers by May [2006] from ecology and design perspective. These three dimensions are then described in four categories: ecological, hydrological, social and spatial. These categories do not exist independently from each other, but each have their influence on each of the other categories. An overview of the three dimensions of connectivity is given in table 2.1. Ecological connectivity includes species migration, material flows and interactions with adjacent ecological systems. Hydrological connectivity encompasses aspects such as riparian interactions and groundwater exchanges. Social connectivity focuses on the human interactions with the urban river and spatial connectivity examines how riverbanks allow access for humans and ecosystems.
The principle open amenity of urban river spaces reflects their capacity to support both ecological processes and urban dynamics. Forgaci [2018] categorizes these under the themes of

	Spatial components of the urban river corridor under the theme of open space amenity
Ecological and water space	Wetlands and floodable areas for water storage capacity, water space defined by cross-section (flow capacity), length and configuration (sinuosity), ecotones as spaces of ecological transition and interaction between land and water, Green corridors and patches along the corridor to accomodate ecological processes.
Public space	Promenades as public spaces designed for the river, Embankments designed to allow access to water, a diverse set of amenities at grade to support the public space of the river, parks and green spaces to provide shade and a pleasant setting for recreational and leisure activities, places of belvedere to improve the visibility of and in the river space.

Table 2.2.: Synthesis of spatial components of the URC under the theme of open space amenity. Retrieved from Forgaci [2018] citing Prominski et al. [2023], Stevens (2009) and Gordon (1996)

ecological and water spaces, and public space (see table 2.2). This dual focus on ecological and public space for spatial capacity ensures that typologies can address both environmental sustainability and urban livability. The principles of integration and multi-scalarity emphasize the importance of understanding the river space as part of a larger urban and ecological system. A typology cannot only be site-specific, but has to be able to adapt to regional and global contexts.

The key properties discussed in this section give insights into important characteristics of the river space for multiple dimensions, and aids the decision making for the fundamental properties used further in this thesis.

3. Methodology

The previous two chapters investigated the needs for analysis of the urban river space and this aids the motivation for the proposed delineation and typology methods discussed in the current chapter. First, a small summary of these two chapters is given. Then, the research design and conceptual framework are discussed, and the proposed delineation methods are described in more detail. After this, methods used in the typology development are given, explaining cross-section segmentation in more detail, and specifying the clustering algorithm and evaluation.

3.1. Summary of related work

Chapter 1 motivated the analysis of the urban river space, indicating the needs for the development of an urban river space delineation and typology. Chapter 2 argued for the choices that will be described in this chapter. The previously mentioned existing 2D river space delineation is based on the first building line. Potential other delineation method were described in section 2.1, as visible space and floodable space. Section 2.2 discussed some existing river space typologies, and identified the need for a segmentation method. As the river space is often visualized as cross-section segment in existing river space analysis, a segmentation method based on these cross-sections is used. Characteristics of the urban river space were discussed in section 2.3 in order to aid the choice of fundamental indicators used in the typology method proposed in this chapter.

3.2. Research design

The research conducted in this thesis is divided into two parts, the typology development and the semantic delineation. The metric-based typology of the river space and corridor is divided into cross-sections along the river, and the longitudinal space as a whole.

1. A typology is derived from analysis of cross-sectional segments along urban rivers in the Netherlands, based on geometric, flood, biological and open space properties of the segments.

The delineation of the river space, consists of multiple parts:

1. The visible space around the river
2. The space around the river bound by the first line of buildings
3. The space around the river vulnerable to flood

The research adopts a primary exploratory approach, addressing the under explored domain of river space typology, and specifically the data-driven approach of river space typology. Existing studies often provide descriptive accounts of these spaces but lack systematic classification, particularly through quantitative analysis grounded in systematic observation. The goal is to discover pattern and relationships in the spatial properties of river spaces and use these findings to create a typology.
This thesis focuses on quantitative analysis of spatial data to identify, segment and classify urban river spaces. Metrics like slopes, land use ratios and visible space are measurable and form the foundation of the typology. Quantitative methods ensure objectivity in measuring spatial metrics. Exploratory research helps identify patterns or trends across scales.

A typology is developed by applying a clustering algorithm to the collected data. A qualitative interpretation of the clusters is then necessary to develop the actual river space typology.

3.3. Conceptual framework

As mentioned before, an unlimited amount of characteristics can be given of the urban river space. Forgaci [2018] gives us a comprehensive overview of river space characteristics. The use cases in 1.2.1 provide an understanding on the importance of characteristics for different urban planning applications. Indicators that are applicable on any river space are needed. The following aspects are taken into account for selecting parameters

- Suitability for cross-case comparison. avoiding overly specific constraints.
- The ability of the parameter to reflect key properties of the urban river space effectively.
- The incorporation of elevation (3D/2.5D) data to add detail to existing methods and reveal information of the space that cannot be researched using only 2D data.
- Data availability, to ensure the practical applicability of the parameters.

One thing all river spaces have in common, and is relevant to all applications, is the fundamental spatial characteristics of the space. This involves elevation and slopes, the river width, and the defined delineated spaces. Additionally, biological components are central to the river space as they are often the main areas of green infrastructure. Similarly, flood indicators are inherently fundamental to the river space due to their nearby river. For the same argument given for the choice of delineating the visible space around the river, visibility indicators are of fundamental importance to the river space type. Lastly, general land use information will provide semantic information on the segments to be able to make sense of what is there following spatial characteristics. Therefore, the following indicators are considered fundamental in this thesis.

- The spatial characteristics of the river space
- Flooding potential considerations
- The interplay between built environment and biological components, the artificial vs. the natural environment.
- The visibility of the river from the river space

3.3.1. Delineation and typology

The delineation of the urban river space is fundamental to the framework, focusing on defining the study boundaries based on key physical markers. For example, the approach by Forgaci [2018] is adopted, delineating the river space by the first line of buildings. Furthermore, the limitations of this approach allow other methods to compliment and strengthen this definition. Therefore, a delineation based on visible space and on flood susceptibility is developed.
Typology is used to categorize spatial configurations and attributes of the river space, such as elevation differences and visibility metrics. This aids the comparison and distinguish different river spaces and understanding their unique features. The parameters are chosen as follows:

- **Geometric properties of the segment**: The segment's shape, dimensions and surface features are essential indicators of spatial capacity and potential use. Elevation differences are critical for assessing flood control and accessibility. Stark height differences disturb accessibility, while gentle slopes allow movement. This research uses the DTM and DSM to compute geometric properties.
- **Visible area**: Visibility is crucial for understanding the connectivity of the space and describing the city-river relationship. Analyzing visible areas, including the impact of obstacles like buildings and trees, highlights how public spaces and visual integration influence city-river interactions. This research uses the DSM to determine visible space.
- **Semantic information and vegetation**: The enhancement of the spatial analysis by semantic information regarding land cover and vegetation allows for a nuanced understanding of spatial characteristics. Spatial characteristics can be determined, but without semantic information not much can be said about the role of these characteristics within the urban river space. This research uses the following categories: vegetation, paved areas, roads, bike paths, pedestrian areas, buildings, and embankment.
- **Flood control**: Flood control is a critical component, as the urban river space is a vulnerable space within the city to extreme weather events. This research uses the 100-year flood data and impervious surfaces to represent the flood vulnerability of the space.

3.4. Delineation

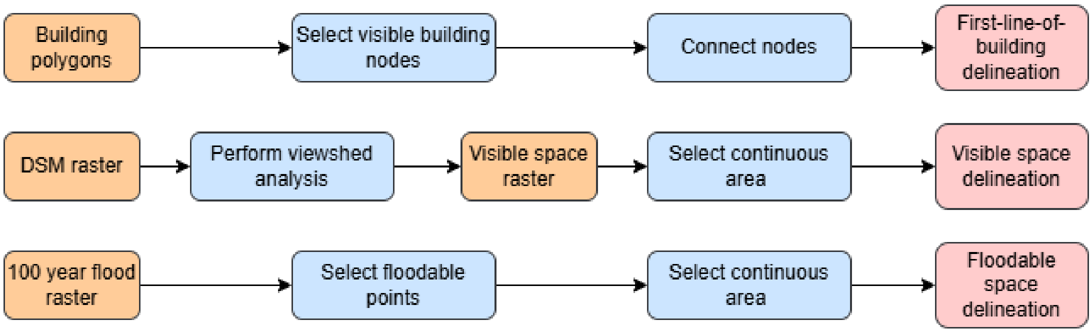


Figure 3.1.: Proposed delineation methods

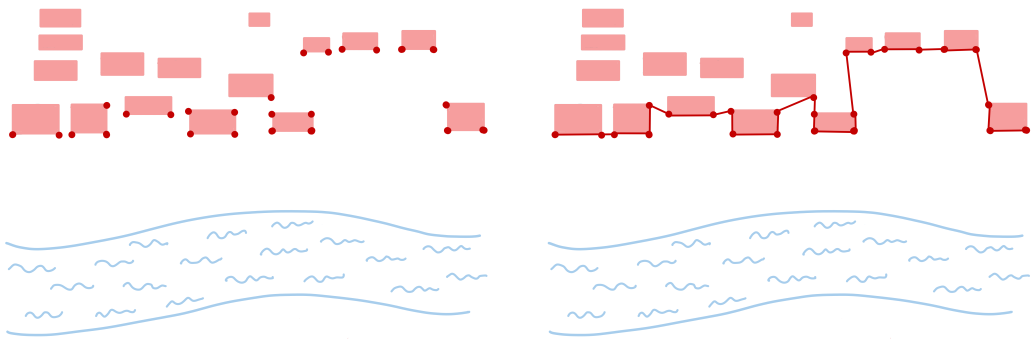
The delineation of urban river spaces is a crucial step in understanding their spatial characteristics and potential. The existing method described in section 2.1.1, defines the river space by the first line of buildings. This approach does not fully consider the spatial characteristics of the area, such as visibility or elevation differences, which can be crucial to gain a functional understanding of the space.
An alternative approach involves using visible space as a key indicator to delineate areas with a visual connection to the river. This area can lay in the river space delineated by the first line of buildings, but could also extend past this line if elevation differences in terrain or if building height differences permit a visual connection. Incorporating visibility into delineation provides a richer understanding of how urban design and topography interact with the river space. Flooding is an inherent risk to the river space. Therefore, delineating areas that are susceptible to flooding next to the urban river is of value. The three methods are summarized in figure 3.1, specifying steps taken to achieve the delineations, and will be discussed in more detail in the following sections.

3.4.1. First line of buildings

The river space delineated by Forgaci [2018] is defined by the first line of buildings, but a structured approach to extracting this definition from spatial data is not provided. Therefore, this thesis proposes a structured approach, based on visible building nodes. Visible building nodes are extracted by creating a visibility graph from midpoints on the river line to these building nodes, see figure 3.2. Visible edges are extracted, and such edges that still fall behind another edge are removed. Nodes are ordered and connected. Input for the method is the building footprints and the river line segment.

3.4.2. Visible space

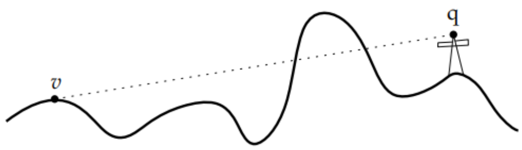
The concept of visible or open space is integral to understanding the connectivity and open space amenity of urban river spaces. This thesis uses the viewshed from raised points on the river line to determine this space. A viewshed uses line-of-sight calculations to generate a map of the view of an area from a specific point, some examples are shown in figure 3.3. Multiple viewsheds along the riverline can be computed to generated the visible space with



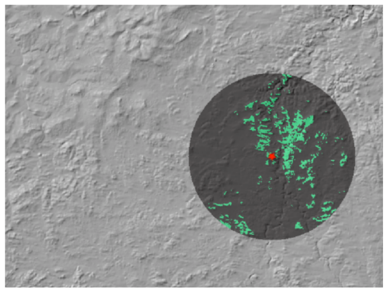
(a) Nodes of buildings (pink) that are visible (in red) from the river are selected

(b) Nodes are connected to create the first line of building

Figure 3.2.: First-line-of-buildings delineation via visible building nodes



(a) Line of Sight calculation



(b) The viewshed of the red viewpoint in which green means visible from this viewpoint. The maximum view distance in dark grey is set to 15 kilometers

Figure 3.3.: Line of Sight and viewshed (Image source: terrainbook)

the river as vantage points. These viewsheds can be combined to cover the whole area. As a viewshed will return 'visible' or 'invisible' for each gridcel in the map, when combining them, the maximum value of all viewsheds at each gridcel can be taken to account for a cel being invisible from one point, but visible from another.
When conducting a viewshed analysis, the choice of elevation data plays a critical role. For instance, the analysis can utilize a DTM, a DSM, or solely building data to identify visible areas around the river. In this thesis, DSM data is selected to determine visible space, as it better reflects the real-world experience of being in the environment. Unlike the DTM, which excludes features like trees and buildings, the DSM provides a more comprehensive representation of the actual visible space. On the other hand, using only building data for the analysis would yield results similar to the delineation by the first line of buildings, though it may also consider structures beyond the initial row.
The maximum extent of the viewsheds is set at a buffered area of 110 meter around the river, which is slightly larger than what is used as segment length in the typology. This ensures data is collected that covers all segments. The river space as visible space in this approach is defined as the reach of visibility. This implies that *holes* in the space are still considered

river space. In order to do this, small separate areas tagged as visible that fall out of the continuous visible space are removed.

3.4.3. Flood delineation

The floodable space in this thesis is bound by the maximum extent of areas susceptible to flood in the 100 year flood depth data. This results in the space around the river deemed unsafe in flood events, as even space that are not deemed floodable but are surrounded by floodable areas are incorporated. These areas need extra attention in urban development due to this vulnerability, and is therefore worth using as river space delineation. This space, just like the first-line-of-building, is bound within the 100 meter buffer to ensure the connection to the river.

3.5. Typology development

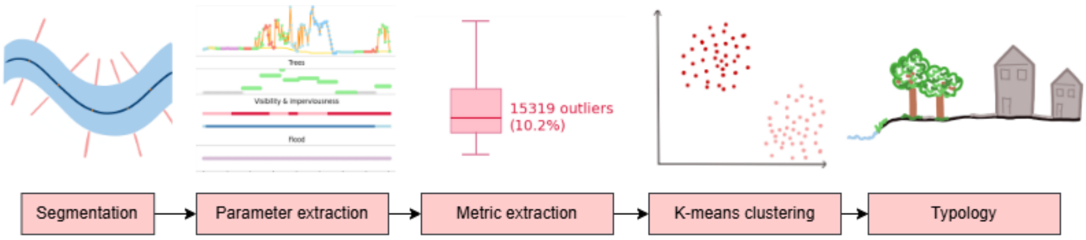


Figure 3.4.: Proposed method for river space typology development

The proposed method for the river space typology development is shown in figure 3.4. It starts with cross-section extraction along the river, continues with data collection as parameter values along the segment and metric extraction per segment following these parameter values. It ends with applying k-means clustering to the metric values and the interpretation of the clusters to establish a river space typology. The following sections will elaborate on these steps.

3.5.1. Cross-section analysis

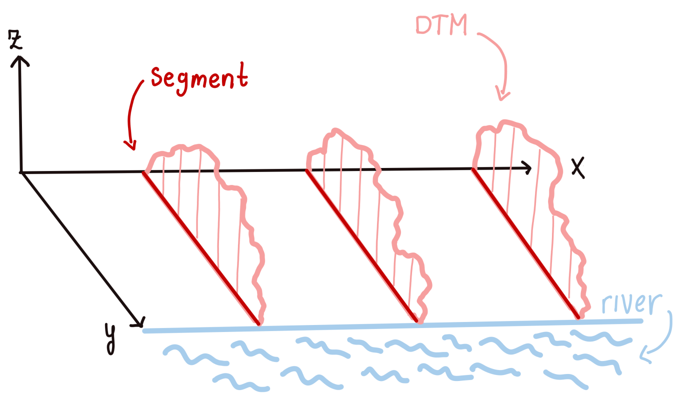


Figure 3.5.: Cross-section line segmentation providing information in the longitudinal direction

As discussed in section 2.2.2, literature often visualizes the river space as a cross-section to show the key properties that characterize a type. A limitation, but also a benefit, of cross-section representation is that they can only provide a simplified, 2D view of complex 3D environments. This results in more interpretable data that focuses on very specific features of the river space. Especially in communicating findings to non-experts such as urban planners or the public, cross-sections are an easily understandable representation of the river's interaction. In a single view, the relationship between different elements of the river space can be made clear. Even though no longitudinal information is provided, using the cross-section may already provide the sufficient detail necessary for the typology without overwhelming the method with unnecessary information from the surrounding 3D environment. If the primary goal is to understand the river in terms of its linear flow and interaction with adjacent areas, cross-sections might be all that's necessary. By analyzing multiple cross-sections at different points along the river, it permits the creation of longitudinal profiles of the river space's behavior. For site-specific design interventions, such as the construction of levees, flood barriers and riverbank stabilization, cross-sections offer the spatial resolution necessary. The main question here to answer is where the additional value of using a 3D model instead of cross-section information lays in the delineation and typology of the river space. Does a full 3D model provide additional value beyond what can be captured by well-placed cross-sections, especially if the goal is to assess typologies within a small or well-defined section of the river space. Some information may be necessary to collect within the 3D model, while others are sufficiently described using only the lateral segment.

This thesis chooses to use the cross-sectional segment as the unit of analysis for river space typology. Cross-sections along the rivers will be selected to analyze. The cross-section is a 2D slice, and thus only provides information in the lateral and vertical direction, as compared to using area patches as segments, which would include the longitudinal direction along the river (figure 3.5). To account for the longitudinal direction, when analyzing the river space of a particular river, the different type of cross-sections present along the river and the order in which the present themselves, can be taken into account. The analysis of the longitudinal direction is outside of the scope of this research, but has to be noted as the river eventually has to be analyzed in its entirety. The methods applied in this thesis can be used for this, but for a full corridor analysis, additional steps would be necessary.

Since cross-section segmentation is a straightforward representation of the space, it is expected to easily scale up to other places. It is not dependent on the Netherlands. In other segmentation methods the issue of longitudinal boundaries arises, which may have to be defined very differently between cities and countries. However, cross-section segmentation does require defining the distance between sections and the length of the sections themselves. A shorter distance between sections provides more segments per river, and can therefore offer more detail in the analysis of river spaces. If segments are very close together, there may be redundant data as the segments would belong to the same river space type. A choice has to be made here. This research uses a distance of 25 meter between segments, and additionally an extra segment for when the river line is bend. This is done to ensure that data is collected in areas, as gaps appear when the river line bends (Figure 3.6).

The length of the segments is also a chosen value. This depends on how the maximum reach of the river space is defined, such that the space still has a clear connection to the river. This research uses a length of 100 meter for each segment. This value is chosen as it is expected to capture enough information about the space to extract a sensible typology. Also, 100 meter can be considered a reasonable distance a person can see on a clear day in the urban environment. This value may be too small or too large, and it open to adjustments in further research. The aim is to not provide an extremely detailed representation of the segment that has every single possible parameter information in it, but to provide a base

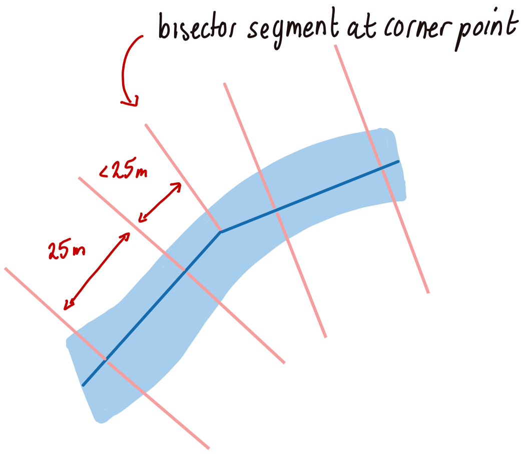


Figure 3.6.: Corner segments

representation, concerning key properties, to be a general reference that can be adapted by adding more information if necessary in particular applications. In effort to avoid overly specific constraints, parameter categories are taken as quite general.
The parameter categories used are described in the beginning of this chapter. From these parameters, numeric metric values will be extracted, as input for river space typology development.

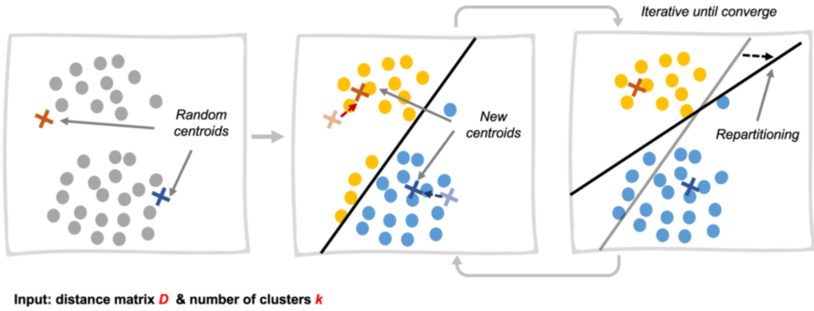


Figure 3.7.: Explanation of the K-means clustering algorithm. Observations are grouped together based on distance to centroid, after which the centroid is re-positioned. (Image source: Gao et al. [2023]

3.5.2. Clustering

A clustering algorithm was applied to achieve a sensible typology of river spaces. Such a data-driven approach for typology development enables large-scale comparative analyses across cities and support urban planning and design practices by providing precise, context-sensitive descriptions of urban forms.
Many different types of clustering algorithms exist, such as partitioning-based algorithm, hierarchical, density-based, grid-based, model-based and novel approaches [Wegmann et al., 2021]. As this thesis does not assess the best clustering algorithm to use for river space typology, a choice has to be made. This research choices to use a simple partitioning based clustering algorithm, namely k-means clustering. An unsupervised algorithm is chosen as the goal is to identify previously undetected patterns in the data, aiding the identification of features useful for categorizing data.

3.5.3. K-means clustering

K-means clustering is a technique which groups unlabeled data into different clusters, and is therefore an unsupervised machine learning algorithm. It groups based on similarity, and works by first picking a given k amount of random centroids, see figure 3.7. Each data point is then assigned to one of these centroids, and together form a cluster. The centroid location is then updated to the average position of the points in the cluster. This step is repeated until the clusters stop changing, and the final clusters are obtained. This illustration immediately shows two difficulties that affect the clustering: the placement of the centroids and the convexity of the data. If data is not well-separated and not convex, k-means clustering can have a hard time finding meaningful clusters. Therefore, the choice of input data and the potential transformation of distributions of input data is of importance. Metrics with a lot of outliers can significantly impact the outcome of the algorithm. Additionally, data requires a scaling transformation due to k-means being a distance based algorithm. This is done in order to avoid any feature becoming predominant in the distance calculation. Different kinds of scalars or transformations can be used, and the choice is dependent on what the distribution of data is like. The aim for the transformation is to reduce the number of outliers as these drag away centroids of clusters and make the clustering algorithm perform worse, and to reduce the skewness within the cluster.

3.5.4. Choosing the amount of clusters

An important step in applying *k*-means clustering is deciding on the value of *k*, the amount of clusters. The choice of amount of clusters is informed by a clustergram. A clustergram is similar to a dendogram. Dendograms can be used in hierarchical cluster analysis to show how clusters are formed, however can be impractical when there is a large number of observations, as each observations requires a node. They are also only suitable for hierarchical clustering, and hierarchical clustering becomes very time consuming when many observations are used. For these reasons, [Schonlau, 2002] proposes the clustergram, which is suitable for a large number of observations, and can also be used to visualize how clusters are formed for a nonhierarchical method such as K-means clustering. Figure 3.8 shows and example with on the x-axis the amount of clusters and on the y-axis the **pca!** (**pca!**) weighted mean. The thickness of the lines shows the amount of observations going into the cluster.

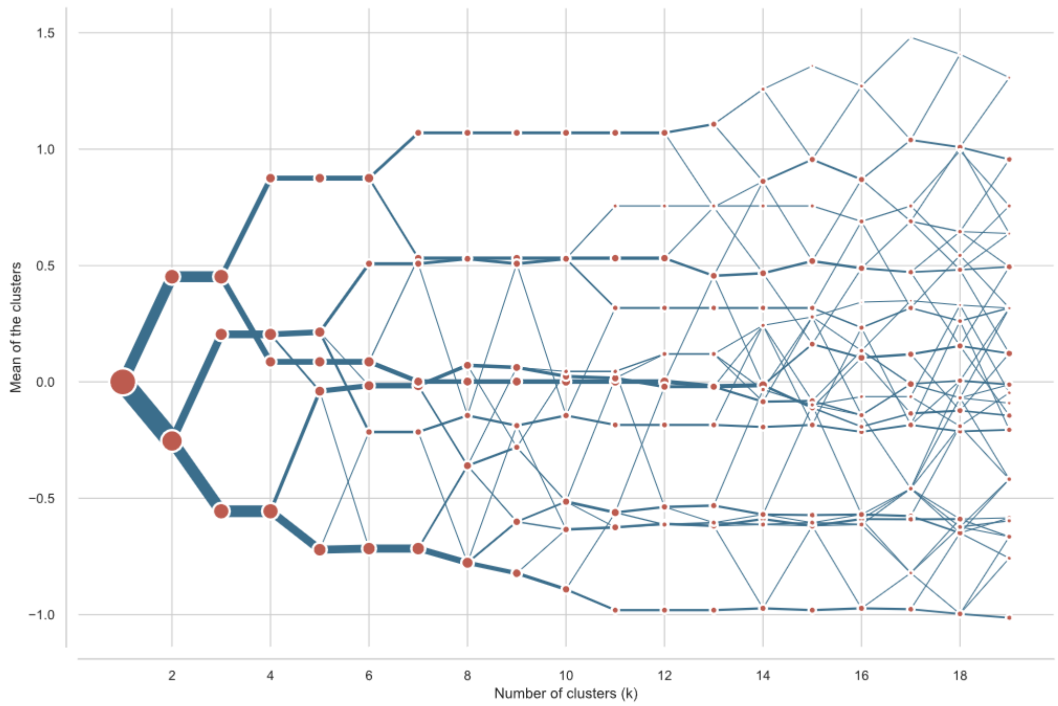


Figure 3.8.: Example of a clustergram. The red dots represent the cluster centers and lines show how observations switch between clusters. The thickness of the lines represent the amount of observations (Image source: Fleischmann [2023])

The clustergram aids the choice of amount of clusters, together with relevant metrics of a goodness of fit such as silhouette score, Calinski-harabasz score and Davies-Bouldin score. The silhouette score [Rousseeuw, 1987] is a measure of similarity of an object to its own cluster compared to the other clusters. The silhouette is based on the comparison of its tightness and separation, and shows if an object lies well within its cluster. A higher score indicates better clustering. The Calinski-Harabasz [Caliński and Harabasz, 1974] is an internal cluster validation index, which measures how similar an object is to its own cluster compared to other clusters. The cohesion is estimated based on the data point - cluster center distances, and the separation is based on the distance of the cluster center to the global centroid. Higher values for the Calinski-Harabasz score indicate dense, well-separated clusters. An optimal value for this score is thus a high value, but we need to choose a solution which gives some sort of abrupt change. If the line is smooth, then then one solution is not necessary better than another. The Davies-Bouldin score is the average similarity measure of each cluster with its most similar cluster. Similarity here is defined as the ratio of within-cluster distances to between-cluster distances. Lower values indicate better clustering.

3.5.5. Interpretation of clusters

After the value of k is decided on and the algorithm is applied, the clusters need to be interpreted and connected to the real-world situations. The distribution of metric values within each cluster is analyzed and compared between clusters, to gain insight on what differentiates a cluster from another cluster. The most representative segment (the segment closest to the center of the cluster), and a selection of randomly sampled segments are analyzed to interpret the clustering results.

4. Implementation

This chapter explains how proposed methods are implemented. It starts with elaborating on the used datasets and tools, and follows with defining the research areas. Afterwards, the process of selecting the segments is explained in detail and the process of adding parameter values to the segments. Next, three algorithms are given for the three delineation methods. Finally, the metrics and their computation are defined and the best amount of clusters is evaluated.

4.1. Data

In order to delineate the river space and create a typology, certain datasets are needed. The following datasets were used in this research, and details are given in table 4.1.

- To determine urban areas in the Netherlands, the population distribution map from Centraal Bureau voor de Statistiek (Central Bureau of Statistics) (CBS) was used. This is the **population distribution** dataset, Infrastructure for Spatial Information in Europe (INSPIRE) harmonized, containing the number of inhabitants per municipality, NUTS2 (basic regions) and $1km^2$ grid according to the European coordinate system ETRS89:LAEA.
- To obtain river line data, **Open Street Map (OSM)** data was used, which is a free, open-source, collaborative mapping database that provides geographic information, including streets, buildings, natural features, and points of interest. It is created and maintained by a global community of volunteers who contribute and edit map data. OSM data can be accessed, modified and shared without restrictions under the Open Database License (ODbL).
- For landuse information and river polygons, The **Basisregistratie Grootschalige Topografie (Basic Registration Large-scale Topography) (BGT)** data was used. This is a Dutch national database containing topographic information. It provides detailed geographic data about the Netherlands, including elements such as streets, buildings, land use, water, vegetation, infrastructure, and other topographical features. The BGT is managed by the Dutch Kadaster (National Land Registry) and is regularly updated to ensure accuracy. The data is accurate up to 20cm. In this thesis, the BGT dataset is used to collect landuse semantics for each segment. The dataset can be download in three formats: CityGML, GMLLight, or StuF-Geo. CityGML is the standard format with most options. GMLLight is a lighter and simplified GML-format with less options. Both of these contain expired/historic objects. CityGml contains more information, for example, has both a point and a polygon layer for the buildings, with the points containing extra information. The GMLLight layer only contains the polygon building footprints. Standaard Uitwisseling Formaat (Standard Exchange Format) (StuF) is a set

Source	Type	Format	Resolution	Purpose of processing	Access
CBS	Population distribution	GeoTIDD	$1km^2$	Determine urban areas in the Netherlands	Open via PDOK
AHN4	DTM raster 2020-2022	GeoTIFF	0.5m and 5m	Add elevation of terrain to segments	Open via AHN
AHN4	DSM raster 2020-2022	GeoTIFF	0.5m and 5m	Add elevation of surface to segments and input for visibility analysis	Open via AHN
OSM	Open street map `LineString` Vector data	OSM XML, OSM JSON	-	River lines for research area selection	Open via OSM
Copernicus	Imperviousness density 2018 raster	GeoTIFF	10m	Add imperviousness values to segments	Open via Copernicus (Account needed)
BGT	Landuse Vector data	citygml & gmllight (.gml), stufgeo	accurate up to 20cm	Add landuse semantics to each point on the cross-section. Water polygons for research area selection	Open via PDOK
LIWO via Klimaat-effect-atlas	100 year flood raster data	GeoTIFF	25m	Add flooddepth value to segments. Floodable space delineation	Open via request at klimaateffectatlas
RIVM	Tree raster data 2022	GeoTIFF	10m	Add percentage of trees to segments	Open via Nationaal Georegister
Hogeschool Amsterdam	Neighborhood typology vector polygon data	GPKG	-	Evaluate developed typology	Open via klimaateffectatlas

Table 4.1.: Details of used datasets

of agreements on the exchange of data between applications in the municipal field. The StufGeo format does not contain the data needed. GMLLight format is sufficient for the purposes of this thesis.

- The **Actueel Hoogtebestand Nederland (Actual Height model of the Netherlands) (AHN)** dataset is the as point cloud in the form of open data. It is available as .laz file. Besides the point cloud itself, they also provide a DTM and DSM model in GeoTIFF format as open data via their website. A QGIS plugin for PDOK exists, such that the data can immediately be loaded into the Geographical Information System (GIS) software.

- The **imperviousness density index** provides at pan-European level in the spatial resolution of 10 meter and 100 meter. It represents surface sealing as a percentage, ranging from 0 percent (completely permeable) to 100 percent (completely impervious), based on the reference year 2018. Via Copernicus Land Monitoring Service. Copernicus is part of the European Union's Space Program, offering information services deriving from Sentinel satellite data.

- The **100 year flood data** is retrieved from the Landelijk Informatiesysteem Water en Overstromingen (National Water and Flood Information System) (LIWO) (website) via Klimaateffectatlas. LIWO contains map layers for professionals who are largely concerned with flooding in the Netherlands. The LIWO is a product of the Watermanagementcentrum Nederland (Water management centre of the Netherlands) (WMCN).
- The 2022 dataset **trees in the Netherlands** by the Rijksinstituut voor Volksgezondheid en Milieu (National Institute for Public Health and the Environment) (RIVM), available via the **Nationaal Georegister** is a 10 meter resolution raster map of the Netherlands indicating the percentage of all trees higher than 2.5 meter in a gridcel.
- The **neighborhood typology** developed by *Hogeschool Amsterdam* is used in effort to validate the developed typology of river spaces. This is a spatial dataset that categorizes urban neighborhoods based on physical characteristics such as building height, density, land use, and street layout. It provides a systematic classification of different urban forms across Amsterdam, identifying patterns such as compact urban blocks, post-war housing estates, and mixed-use areas. The typology is designed to support urban analysis and planning by offering insight into the structural and functional makeup of various neighborhood types.

4.1.1. Data quality

Many different data sets are used in this thesis. As mentioned, the datasets have different spatial and temporal resolutions. Temporal mismatches can arise when datasets are collected in different years, for example the AHN4 dataset from 2020-2022 and the Copernicus imperviousness dataset from 2018. Spatial differences may make datasets not align, which was noted during implementation with the AHN4 data and BGT data not always aligning. Data is collected in different ways, such as systematic government programs, or via community contributions with varying levels of reliability. Such differences impact consistency and comparability. There are differences in how objects are tagged, with OSM tagging differing from BGT classifications. These are a selection of problems that can arise when combining different datasets. Quality of data and coherence between datasets is of importance.

The AHN dataset offers a description of the data quality on their website. The accuracy of the height points do not have more than 5 centimeter standard deviation and not more than 5 centimeter systematic deviations. The AHN4 dataset was collected between 2020 and 2022. This thesis uses the rasters derived from the point cloud. A value for each 50x50 centimeter cell is computed using Squared Inverse Distance Weighting. The AHN is a collaboration between the *waterschappen*, provinces and *Rijkswaterstaat* and is made open by the Dutch government to stimulate reuse of available data.

The accuracy of the OSM depends on what data is used of the OSM dataset. OSM can be edited by anyone and therefore may include inaccuracies. To retrieve data from the OSM, tags can be used, but as a user you are in this case dependent on the tagging standards and quality. This thesis uses the river lines from this dataset. There are some alignment issues between OSM river lines and official datasets such as the BGT. Section 4.3 shows an example of a river line that falls on land in the BGT dataset. The BGT dataset is used in this instance to 'clean' the retrieved OSM data. The BGT is an official open dataset offered by the Dutch government via *Kadaster*, and its widely used within the Netherlands as the use of the dataset is mandatory for governmental institutions.

The imperviousness data is retrieved from Copernicus, has a position accuracy of less than half a pixel, and a thematic accuracy of over 90 percent. The data is updated every 3 years, with the latest update being from 2018.

The used tree data is made available by the governmental organisation RIVM and is a product of the AHN4 dataset, the buildings from the Basisregistratie Adressen en Gebouwen (Basic Registration of Addresses and Buildings) (BAG) dataset, and the infrared images of the Netherlands. The RIVM is in charge of the dataset and metadata. The `readme` file that comes with the dataset states that they can not guarantee the abscence of faults in the dataset, and does not offer accuracy or computation specifications. As this is a derived dataset, quality depends on the quality of the previously mentioned datasets used to generate this raster.
The flood data is retrieved from the *Landelijk Informatiesysteem Water en Overstromingen* (National Informationsystem Water and Flood) and is a product of the WMCN which is part of the governmental institution *Rijkswaterstaat*.
These derived datasets offered by governmental institutions do not have quality specifications provided at the websites where they can be retrieved.

To combat problems associated with different datasets, the OSM data is checked with the BGT dataset. Even though data is collected in different years, it is assumed that most data aligns as the difference between datasets is not more than a couple of years (for datasets where this is specified). The most recent dataset available is always taken. Datasets are taken from official institutions to ensure sufficient quality (except for OSM). As data collection is out of scope, this thesis relies on available data and thus using datasets which are also used by Dutch government results in sufficient data quality.

4.1.2. Identified problems in used data

Even though the datasets come from official institutions, there are two identified problems that needed to be solved during implementation.
A problem which arise when adding flood depth data to the sections was some very high values which skewed the end results. This only occurred in one section in the flood depth map, as shown in 4.1. Sections that contained flood depth values higher than 20 meter were discarded of.

Another concern was that in the BGT dataset, columns would have slightly different names in different parts of the dataset when downloading pieces of it, or sometimes have the column in Dutch and sometimes in English. Each `FeatureType` required the extraction of data in a specific column, and the column names in table 4.2 were used. Similarly, to remove temporal data, columns *objectEindTijd* and *terminationDate* were used. Additionally, if an object had a *eindRegistratie*, it was removed. For a couple of downloaded layers, columns *objectEindTijd* and *terminationDate* were not present. In those cases, data was only filtered using *eindRegistratie*.

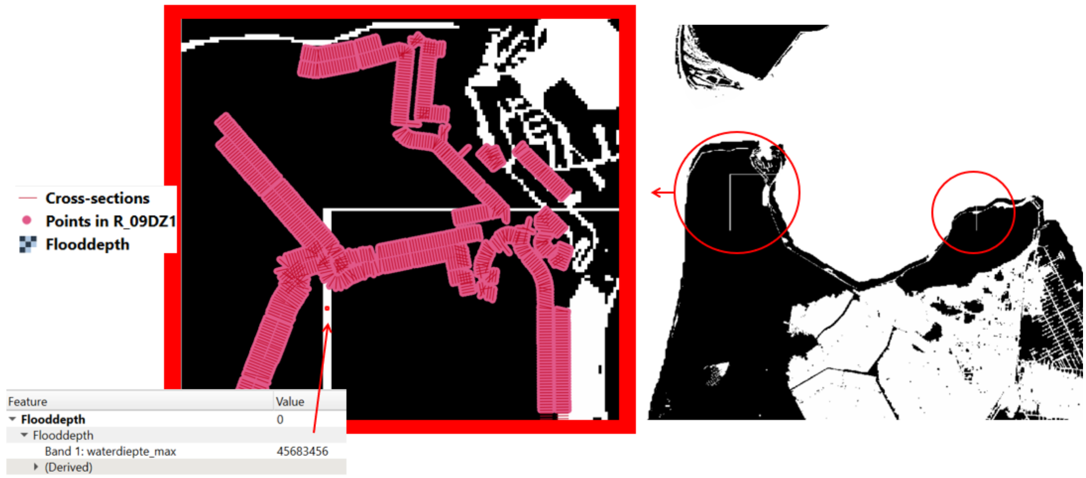


Figure 4.1.: Inaccuracies in the flood depth dataset. Two areas are shown on the right image where values are exceptionally high (red circles).

FeatureType	Columns
Onbegroeidterreindeel, begroeidterreindeel and ondersteunendwegdeel	*bgt-fysiekVoorkomen*, *bgt_fysiekVoorkomen* and *class*
Wegdeel	*bgt-functie*, *bgt_functie* and *function*
Waterdeel and ondersteunendwaterdeel	*bgt-type*, *bgt_type* and *class*

Table 4.2.: `FeatureTypes` and used columns in the BGT dataset

4.2. Tools

The algorithm for metric extraction and river space delineation is implemented in Python programming language. Python offers a great number of open-source libraries that can be used for spatial analysis and is implemented in the GIS software QGIS. The following libraries were used in this thesis:

- **Geopandas**: Geopandas is a open source project for geo-spatial data processing in python. It uses shapely for geometric operations and Geopandas objects can thus act on shapely geometry objects, matplotlib for plotting, and pyogrio for file access. It is designed to make working with geo-spatial data in python easier, and can do operations in python that would otherwise require a spatial database. It adds support for geographic data to pandas objects and offers subclasses of pandas.Series and pandas.Dataframe as Geoseries and GeoDataFrame respectively. geodataframes for data processing. The GeoDataFrame, a tabular data structure containing at least one GeoSeries column storing geometry, was used in the code to store and process data, together with the reading and writing to files functions that can take almost any vector-based spatial data format. Function .sindex was used to add spatial index to geodataframes to speed up processing. Function .sjoin was used to spatially join geodataframes.
- **Shapely**: A python package that uses functions from the GEOS library, which is a

C/C++ library for computational geometry specifically for algorithms used in GIS software. Classes primarily used were *Point*, *LineString* and *Polygon* to which methods such as *object.bounds* and *object.area* could be applied.

- **SAGA GIS** is a geographical information system computer program used to edit spatial data. Modules from SAGA GIS are implemented within the QGIS environment. This thesis uses the module visibility analysis, which takes as input points and elevation grid, and outputs the viewshed as visible cells in the grid from the input points.
- **Rasterio**: A python package for working with raster data, providing a pythonic wrapping around Geospatial Data Abstraction Library (GDAL). GDAL is a software library for reading and writing raster and vector dataformats and forms the basis of most software for processing geo-spatial data.
- **NetworkX** is a python package for the exploration and analysis of networks [Hagberg et al., 2008]. This package is used to build a network for building data and midpoints, to construct a visibility graph for building delineation. Documentation is available via **Networkx**.
- **scikit-learn**: An open source python library containing simple and efficient tools for predictive data analysis built on NumPy, SciPy and matplotlib. It offers a range of machine learning algorithms, and offers functions for the following clustering algorithms: K-means, affinity propagation, Mean Shift, Spectral clustering, Hierarchical clustering, DBSCAN, HDBSCAN, OPTICS, BIRCH. This thesis uses the Yeo-Johnson transform, K-means clustering, Silhouette score, Calinski-Harabasz Score and Davies-Bouldin score modules.

GIS software in addition to the python scripts were used to perform visibility analysis, namely:

- **QGIS**: An open-source geographical information system supporting Windows, macOS and Linux. The software allows users to edit, analyze, and visualize geographical data. Within QGIS, the python console plugin was used to prepare data files, and perform visibility analysis. Visualizations of input data and derived data were done in QGIS. Functions that were used within the QGIS environment: Filtering of layers by expression, reproject to EPSG:28992, clip vector by Mask Layer, intersection, dissolve polygon, buffer and from Multipart to Single part.

4.3. Research areas

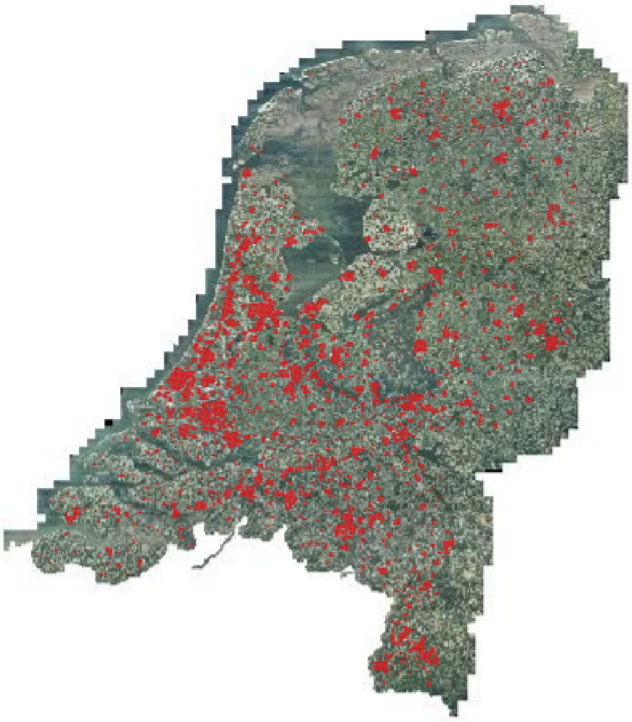


Figure 4.2.: Urban areas as grid cells with a population density > 1000 per $1km^2$ in red. Population density data via PDOK from CBS. Background is Orthophoto 25cm via PDOK.

This thesis investigates the urban river space. To be able to select research areas, the urban area must be defined. This research considers the urban area as an area with a population density of at least 1000 people per $1km^2$. In principle, an urban area is a settlement with a relatively high density, which creates a certain population critical mass and centrality characteristics to a city. Depending on the country, the definition of urban may differ.
An in-depth explanation of the distinction between rural and urban is not as relevant for this particular research, as a method for delineation and river space typology is proposed, which utilizes a definition of urban space, but is not about defining what urban space is. Therefore, the distinction here is given as a population density of 1000 people per $1km^2$, and only the rivers that fall into these areas are taken into account. A value of 1000 is chosen as this ensures the area is urbanized, but may not include all urban space in the Netherlands.

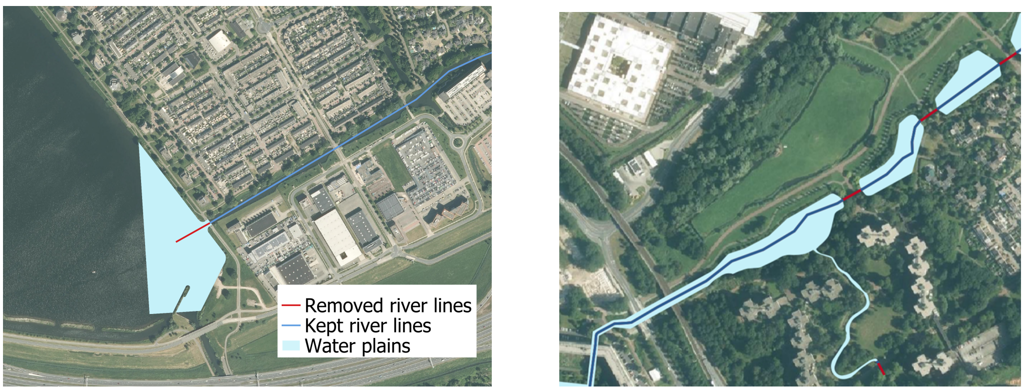
The population density grid is retrieved from PDOK provided by CBS with reference year 2012 with resolution $1km^2$, and filtered by `"obsvalue" > 1000` to represent urban areas, see figure 4.2

To select the rivers, OSM and BGT data is used. OSM data for the river midline data as `key: waterway` with tags `canal` and `river` were retrieved via theOSM overpass Application Programming Interface (API) for the whole of the Netherlands. Ignored were tags `ditch`,

	Free flow		Pipe flow	
Artificial	Open air	culvert	flooded tunnel	pipeline
	canal (useful water)		pressurised	
	ditch			
	drain (superfluous water)			
Natural	Open air	cave		
	river			
	stream			
	tidal_channel			

Figure 4.3.: OSM waterway types. Image retrieved from OSM wiki

`drain`, `stream` and `tidal channel`, see figure 4.3. This dataset was clipped using the 'Clip Vector by Mask Layer' provided by GDAL in QGIS. The OSM file contained duplicates. To identify duplicates, a column `length` was added to the river layer. Duplicates were removed from the layer using 'Delete duplicates by attribute' selecting columns `osm-id` and `length`. Further inspection of the retrieved OSM data revealed that these river lines sometimes laid in waterplains. As these would cause issues in segment selection, most of these were removed using BGT data *waterdelen* (water plains). Not all waterplains were correctly tagged as such, and therefore some are kept. This layer was imported in QGIS, and temporal data was removed by only keeping items with `ObjectEind = NULL`. The layer was filtered and only items with `"bgt-type" = 'watervlakte'` were kept. Then harbours were removed from the file, as these are kept in our research areas. The layer was filtered `"plus-type" != 'haven'`. This resulted in a layer that contained all water plains in our urban areas. The OSM river line layer was clipped using 'Clip Vector by Mask Layer' with as mask layer these water plains. This removed any OSM river line part that was laying in a water plain.



(a) River lines that lay in water polygons defined as 'water plains' are removed

(b) Red riverlines are removed as they don't fall into a water polygon

Figure 4.4.: Cleaning riverlines

The last cleanup step for the riverline data was to remove the parts of the riverline that did not lay in water. There is some kind of structure above the river, and thus the perpendicular area will not be considered valid river space. No embankment point would be able to be found. This is done in QGIS using 'intersection' between the riverline layer and the water polygon layer. Examples for this are shown in figure 4.4b. Finally, it was noticed that a river consisted of a lot of separate lines, or a river consisted of multiple lines that lay far apart due to removing the in between sections that lay in non-urban areas, which made certain processing steps more challenging, so the layer was 'dissolved', then split again into single parts, and then a column was added that contained a new unique ID number for each separate riverline. Additionally, columns not needed for processing were removed to reduce file size.

Figure 4.5.: Separate river linestrings of the same riverlines were merged together.

4.4. Segmentation

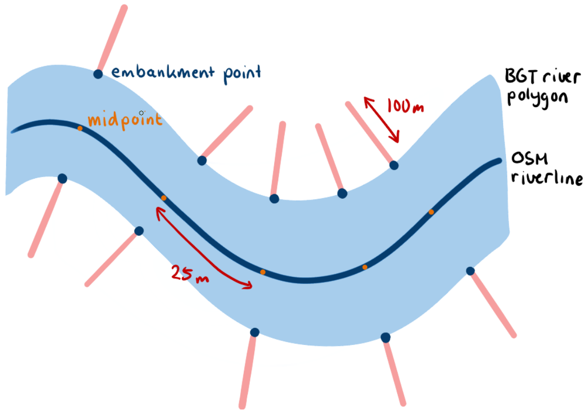


Figure 4.6.: Overview of how segments are selected

Cross-sectional segments are used as input sections for the typology. Figure 4.6 shows how segments are selected. The process to compute segments consists of four steps. I refer to the first step as the pre-process. This consists of creating initial cross sections, crossing a midpoint on the riverline and laying perpendicular to the riverline. The width of this section needs to be long enough to reach the embankment of the river. The next step is to retrieve these embankment points. This is done by intersecting the half cross-sections with river polygons. The third step is the extraction of our final segments, and removing ones deemed invalid. Finally, points on the segment 0.5 meters apart are computed. These points are used to collect data in the section, and a value of 0.5 is used as this value is the same as the raster dataset used with the highest resolution. The following sections explain these steps in detail, and an description of all the steps and how they are done is given by algorithm 4.1.

4.4.1. Pre-process

The goal is to extract `LineString` objects with an embankment point as starting point, extending 100 meters, that lay perpendicular to the river line, every 25 meters along the riverline. To be able to retrieve the sections, embankment points need to be found. This is done using 'initial' cross-sections. These are extracted from the riverline, using points on the riverline 25 meters apart, and with a width of 600 meter. 600 meter is chosen as this width would be able to capture the widest river in the dataset.
Additionally, each corner point in the river gets a half cross-sections, as the bisector of the outward angle to account for gaps (figure 4.7).
These initial cross-sections are split in half at the midpoint and cleaned up. Initial cross-sections for which the midpoint lays on a bridge are removed. This is done using BGT dataset's layer '*overbruggingsdelen*' (bridges).

Algorithm 4.1: Segmentation($\mathcal{F}, \mathcal{S} \leftarrow \mathcal{L}, \mathcal{P}$)

Input: Cleaned riverlines as `LineString` $\mathcal{L}$, cleaned river extends as `Polygon` $\mathcal{P}$, bridges, non-urban areas.

Output: Perpendicular river space segments as `LineString` $\mathcal{F}$ and `Point` $\mathcal{S}$ along line 0.5m apart, extending 100m from embankment points

```
for l in L do
    // Getting initial sections
    M ← Get midpoints as points along l 25m apart;
    C ← Get corners as points of l with angle < 170
    for m in M do
        I ← append initial two perpendicular half cross-sections with width 300m
          and starting point m;
    for c in C do
        I ← append initial section extending 300m outwards from corner point c;
    for i in I do
        if i intersects bridge then
            remove
        // Embankment points
        E ← append embankment point e as first intersection of i with P;
    for e in E do
        W ← append riverwidth as LineString from m to e;
        for w in W do
            if w intersects L\l or 3 * σ > len(w) > 3 * σ then
                remove
            for w intersection W\w do
                if intersection < 3 then
                    keep w closest to mean direction
                else
                    directions = line direction for line in intersection;
                    deviations = angle difference(angle, median angle) for line
                      in directions;
                    if line in deviations > std from median then
                        remove line from W
    // Final sections
    F ← Create final segments extending 100 m from embankment points;
    for f in F do
        if f in non-urban area or f intersects embankment line then
            remove
        S ← Create points along f 0.5 m apart;
return F, S
```

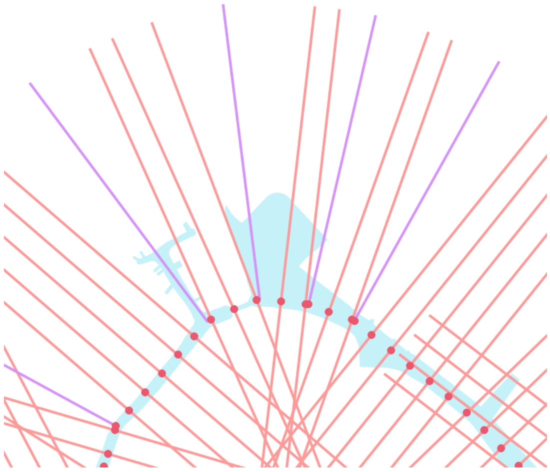


Figure 4.7.: Initial cross-section selection. Corner sections are shown in purple, normal sections in pink.

Figure 4.8.: Midpoints are removed if they intersect with a bridge polygon.

4.4.2. Embankment points

Finding the embankment points is done using intersection points of the initial cross-sections with water polygons of the river. This is done using BGT data, namely the layer 'waterdelen' (water plains), which consist of polygon objects of all water plains and waterways in the Netherlands. To be able to compute embankment points from this file, the BGT data needed to be processed, as it contained temporal data and consisted of many separate polygons, which was undesirable for embankment point computation.
The steps taken to prepare the BGT *waterdelen* (water plains) file and to obtain the water plain polygon input file were as follows: 1) downloaded using the download API for BGT data, retrieving featuretype '*waterdelen*' for polygon bounds of the defined urban areas 2) data was clipped using 'Clip by Mask layer' to the extends of urban areas 3) Data was

filtered by removing objects in the dataset where `"objectEindtijd" != NULL` 4) The layer was dissolved into a single polygon. 5) Small holes in the polygon with an area $< 1000m^2$ were removed. This value is used after manually trying out some smaller and bigger values. A too large value would result in spaces that lay in between rivers being filled in. The value has to be as large as possible to get rid of these holes in the river which would interfere with segment selection.

Embankment points are computed as the first intersection point of each initial half cross-section with the water plain input polygons. A file containing `LineString` objects representing 'river widths' is created. These are the lines from midpoint to embankment point. This file is used to filter out three types of problems that were identified:

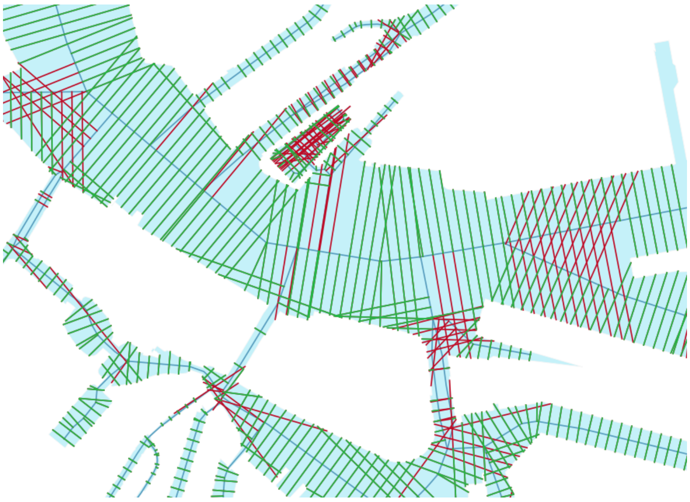
1. Remove river widths that cross another river line, see figure 4.9a.
2. Remove river widths when it's much smaller or larger than the average of river width, using an interval of $3\times$ *standard deviation* with respect to the *mean*, see figure 4.9b.
3. If a river width crosses another river width, the we are probably at some sort of intersection of different rivers, where rivers split up. Here the segment extraction is more difficult. The direction of the river width is computed, and only certain river widths are kept.
    - An intersection graph is created. Intersection groups are computed as all the lines that are somehow connected. So when line A and B are connected, B and C, but not A and C, then the group still includes all three A-B-C.
    - Depending on the direction of the lines in the group, it is decided if a line is kept or not. See figure 4.9c for examples of removed lines.
        - When a group's size < 3, then only the riverwidth with direction closest to the mean is kept, and the other is removed.
        - When a group's size ≥ 3, riverwidths within a standard deviation from the median are kept, others are removed. This is done in effort to exclude lines that lay almost parallel to a river, and in most cases gives the desired result. However, some cases are a bit more complicated, and then it removes them more arbitrarily. This is not considered a problem. These cases have multiple lines crossing each other, which causes some redundancy in data collection, and reducing the amount should not cause significant errors.

When a river width line is removed, the corresponding embankment points are also removed.

4.4.3. Final segments

From the embankment points, the final sections are computed, extending 100 meter away from the river, computed using the direction of the midpoint to the embankment point. Some of the final sections were removed. The goal was to obtain as many 'clean' segments as possible, and to not include certain edge cases. Two types of segments were removed:

1. Segments partially laying in a non-urban area. This happens when the river lays close to the border of the urban area.



(a) Riverwidths crossing another riverline are removed

(b) Relatively large riverwidths are removed

(c) Riverwidth intersections are taken care of

Figure 4.9.: Cleaning riverwidths. Red lines are removed, green lines are kept.

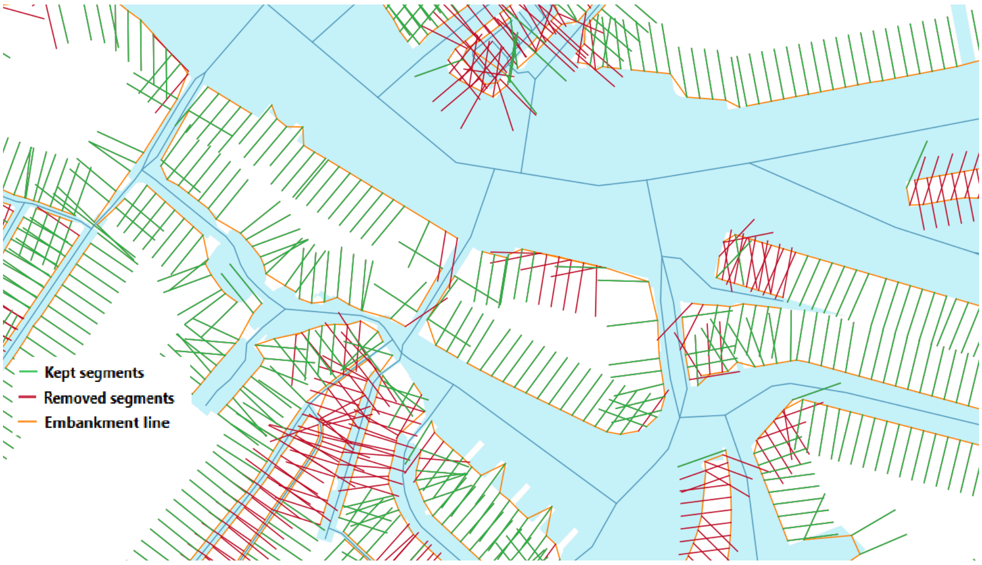
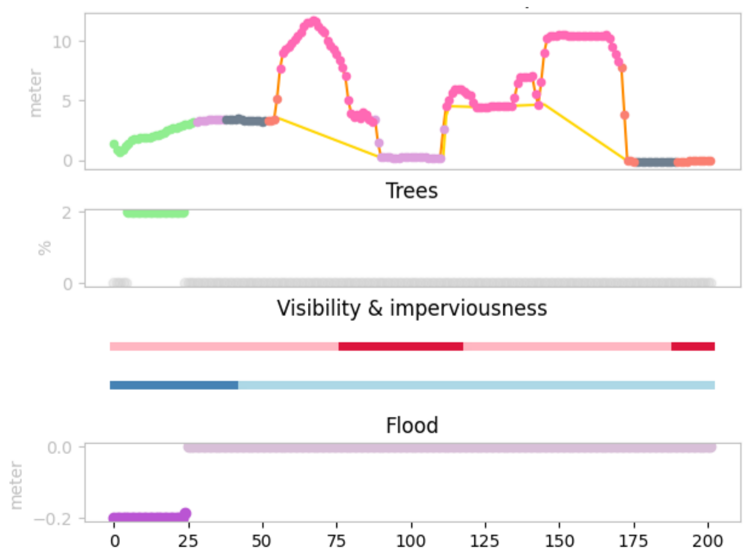


Figure 4.10.: Segments crossing another river's embankment line are removed

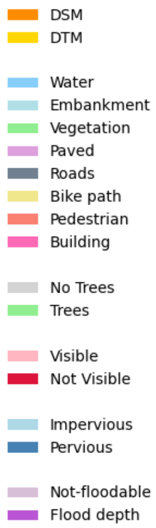
2. Segments crossing another river's embankment line. This implies the segment is shorter than 100m and seen as edge case thus not taken into account, see figure 4.10.

To find the needed embankment line, embankment points for each seperate side of each river were connected to create a file containing `LineString` embankment lines. If a segment intersected with an embankment line of another river, it was removed. As a last step, points are computed as `Point` objects on these segment `LineString` objects, 0.5 meter apart from each other. These points are used to collect data for the segment.

4.5. Parameters



(a) Parameters



(b) Legend

Figure 4.11.: Example of parameter values on a segment

Table 4.3 shows the parameters categories added to each segment. The BGT dataset provides information on landuse. The used layers in this research of this dataset are described in appendix B. These are layers relating to water, vegetation, paved areas including roads, and buildings. This data is added to each point of the segment. Additionally, elevation of terrain, both DTM and DSM, are added from the AHN dataset, see figure 4.11.
The visibility analysis is performed using SAGA GIS visibility analysis module on the AHN DSM with a 5 meter resolution. This raster was clipped to a 100 meter buffer around the water polygons used before to determine embankment points. The analysis is performed from the midpoints on the riverline. Height is defined as the maximum height of the embankment points, and if no embankment points is present, a default of 1.75 meter is used. The visibility analysis is done per separate river, and afterwards merged. If a pixel is visible in any of the overlapping visibility rasters, then it is deemed as visible. Only when the pixel has only 0 values, it is deemed as not visible.
Fundamental flood parameters are defined as imperviousness of ground, and the flood depth in case of 100 year flood. An imperviousness value and a flood depth value are given to each point on the segment. Figure 4.11 visualizes how the parameter values are collected on a segment. The upper plot shows the BGT data per point together with the DSM and DTM lines in meter. The second plot shows tree percentages of the pixel the point intersects with. Visibility and impervious surfaces are shown as horizontal lines in pink-red and lightblue-darkblue respectively. Flood depth in meter is shown in purple in the bottom plot.

Parameter	Description	Source
dtm	Elevation value of point extracted from the DTM	AHN4
dsm	Elevation value of point extracted from the DSM	AHN4
visibility	Visibility value for each point: 1 for visible, 0 for not visible	AHN4: DSM, SAGA GIS visibility analysis
water	Points classified as water	BGT: *waterdeel*
embankment	Points classified as embankment	BGT: *begroeid terreindeel: oever, slootkant*
vegetation	Points classified as silt, forest, heather, bush, greenery, grass, dune, swamp, salty marsh, agricultural	BGT: *begroeid terreindeel: slik, loofbos, gemengd bos, naaldbos, heide, struiken, houtwal, groenvoorziening, grassland overig, duin, moeras, kwelder, fruitteelt, boomteelt, bouwland, grassland agrarisch*
trees	Percentage of trees in intersecting grid cell	
paved	Points classified as paved	BGT: *onbegroeidterreindeel: gesloten verharding, open verharding, parkeervlak, inrit*
yard	Points classified as yard	BGT: *onbegroeidterreindeel: erf, woonerf*. These are areas that belong to a building, often paved, but sometimes contain some green space.
roads	Points classified as road	BGT: *wegdeel: rijbaan, OV-baan, overweg, spoorbaan, baan voor vliegverkeer*
bike	Points classified as bike path	BGT: *wegdeel: fietspad*
pedestrian	Points classified as pedestrian	BGT: *wegdeel:voetpad, voetpad op trap, voetgangersgebied*
building	Points classified as building	BGT: *pand*
imperviousness	Imperviousness value range 0-100	Copernicus
flooddepth	Depth of flood if floodable, Nodata value −9999 if not floodable	LIWO

Table 4.3.: Parameters of segments

4.6. Delineation

The delineation of the river space consists of three parts, as discussed before, and defined here in table 4.4.

Visual connectivity	Bound by viewshed on the DSM with viewpoints at the midpoints of the river and radius extending 110 meters beyond embankment. Defined as maximum reach of visible space.
Building	Bound by the first line of buildings.
Floodable area	Bound by 100 year floodline. Defined as maximum reach.

Table 4.4.: River space delineated areas

The goal for delineating the visible space is to find the area around the river that has

Algorithm 4.2: Visible area($\mathcal{C} \leftarrow \mathcal{P}, \mathcal{E}, \omega$, `dsm`)

Input: Riverlines as `Points` $\mathcal{P}$, embankment $\mathcal{E}$ as `Points` and their `dtm` height, buffer width ω, `dsm` raster
Output: Visible space around river as raster $\mathcal{D}$

```
// Prepare river point file and execute viewshed algorithm
for p_i in P do
    e_1, e_2 ← corresponding e_i in E;
    vheight = maximum of ( e_i['dtm'] ) + 1.75;
    v = SAGA GIS visibility analysis(input: clipped dsm, points: p_i, observer
      elevation: vheight);
    V ← add v
// Create visible area
for v in V do
    C = logical OR(C, v)
return C
```

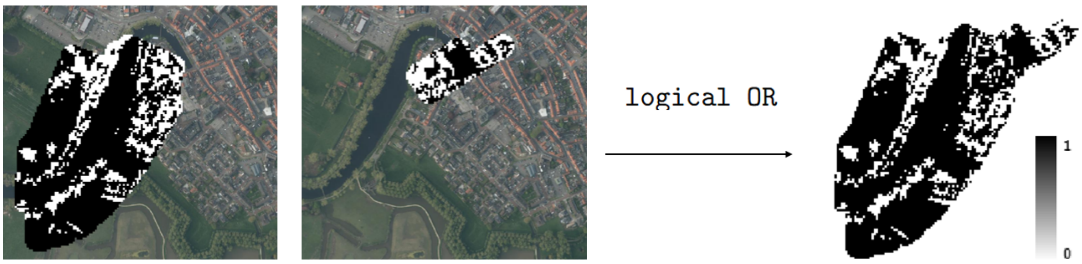
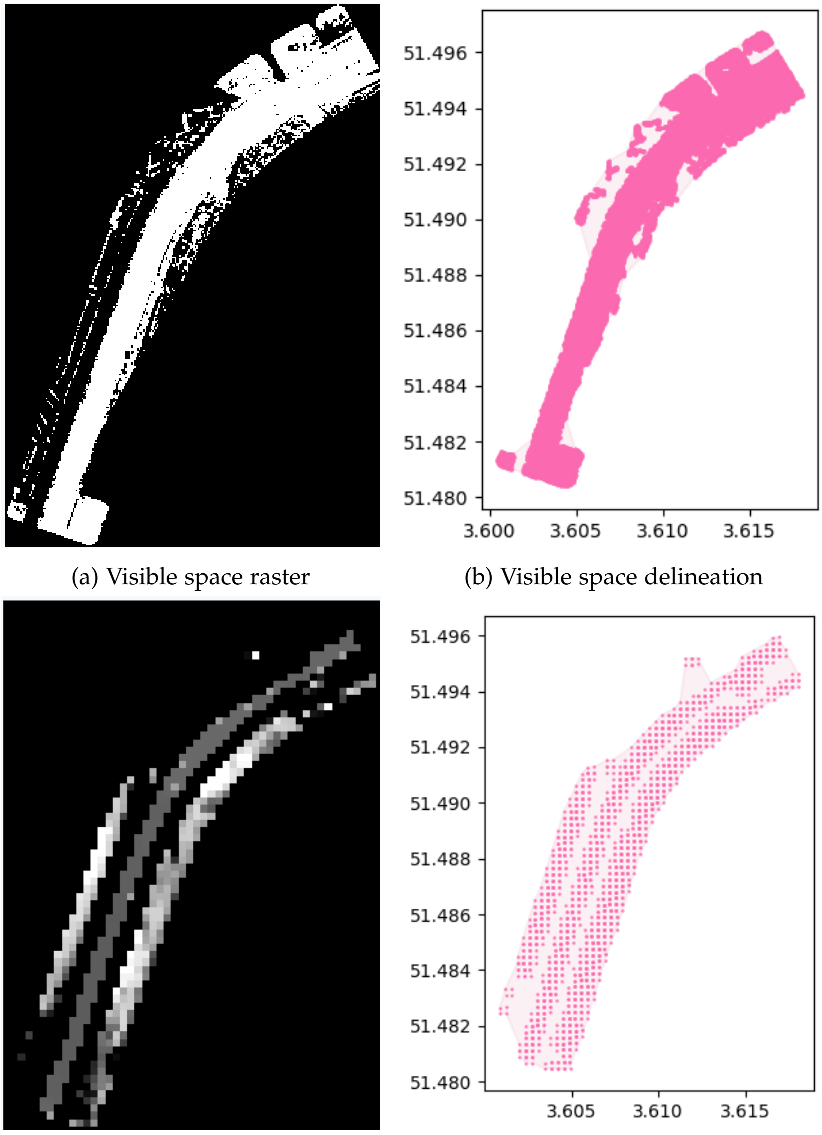


Figure 4.12.: Viewsheds are computed per riverline and combined by identifying each pixel that is visible in any of the viewsheds as visible

such direct interaction with the river. Viewsheds are computed from each midpoint of each cross-section. Certain manually-defined parameters are decided on for this step:

1. Radius: Max extend of the space around a viewpoint in which we compute for visibility. Here this is defined as the distance from the river center point to one of the embankment points (the maximum distance is taken), plus 110 meters.

2. Viewpoint height: Height decided on to compute visibility from. Height is taken as: highest elevation value of embankment (DTM) + 1.75 meters. If no elevation value of the embankment is found, then a default value of 1.75 is used.



(a) Visible space raster (b) Visible space delineation

(c) Flood raster (d) Floodable space delineation

Figure 4.13.: Delineation from raster via alpha shape

To perform the viewshed analysis, the module *Visibility analysis* by SAGA GIS has been used. This function computes the viewshed around a point on an elevation raster map, and gives each gridcel either a visible (=1) or invisible (=0) value.
To delineate the visible space as an area, the viewshed rasters of all the rivers are combined to obtain a single viewshed. As we look for the space that has visible interaction with the river, any pixel in the raster that is visible in at least one of the viewsheds is marked as visible. Steps to derive the visible area raster are described in algorithm 4.2.

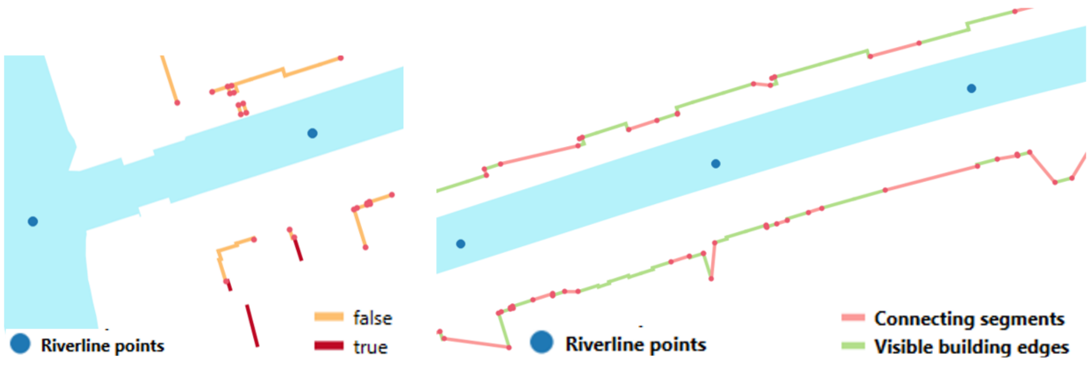
The floodable area is determined by the 100 year flood data within a buffered area around the river. The maximum reach of this floodable area can be determined by computing an alpha shape around this space. The same method is used to determine visible space delineation. As input, the raster file is used and the visible pixels become input for the alpha shape. Small areas that are deemed visible or floodable, but are not connected to the rest of the space, are discarded. This is done grouping visible pixels by their 8-connectivity. If such a group is smaller than a certain threshold (in the algorithm defined by β), then these pixels are not taken into account in the delineation. This implies there are three values to be defined by the user: the buffer value, α value for the alpha-shape, and the minimum amount of pixels. Algorithm 4.3 describes the steps taken to delineate the river space from raster data.

Algorithm 4.3: Riverspace from raster($\mathcal{D} \leftarrow \mathcal{P}$, ω, raster, α, β)

Input: A river as `Polygon` $\mathcal{P}$, buffer width ω, raster file, value α, minimum amount of connected pixels β

Output: Riverspace boundary $\mathcal{D}$ as `Polygon`

$\Omega \leftarrow$ get buffer Polygon with width ω around $\mathcal{P}$;
$\mathcal{A} \leftarrow$ `Clip` raster file by extent Ω;
// Get points
for *cell in* $\mathcal{A}$ **do**
    if *cell is not* `None` **then**
        $\mathcal{C} \leftarrow$ add cell to visible cells
// Discard small areas
$\mathcal{G} \leftarrow$ get 8-connected groups
for g *in* $\mathcal{G}$ **do**
    if $g \geq \beta$ **then**
        $\mathcal{E} \leftarrow$ add corner `Points` of cells
$\mathcal{D} \leftarrow$ `alpha shape` ($\mathcal{E}$, α)
return $\mathcal{D}$



(a) Edges that lay behind another edge are removed (in red, True for 'is hidden')

(b) Visible building edges are connected to form a continuous line

Figure 4.14.: First line of building delineation

Algorithm 4.4: Riverspace building line($D \leftarrow \mathcal{L}, \mathcal{B}$)

Input: A riverline as `Linestring` $\mathcal{L}$, building footprint Vectordata $\mathcal{B}$ per side.
Output: Riverspace boundary line $\mathcal{D}$ as `Linestring`

$\mathcal{N} \leftarrow$ get building nodes
$\mathcal{V} \leftarrow$ Initiate visible points
// Visibility graph
for *n in* $\mathcal{N}$ **do**
  if *n is visible from* $\mathcal{L}$ **then**
    Append n to $\mathcal{V}$
// Remove hidden edges
$\mathcal{E} \leftarrow$ get visible edges
for *midpoint m of e in* $\mathcal{E}$ **do**
  $r \leftarrow$ Shoot ray from m to riverline
  if *r intersects another e* **then**
    Edge is hidden, remove e from $\mathcal{E}$
// Connect visible edges
$\mathcal{G} \leftarrow$ make groups of connected edges of $\mathcal{E}$
for *g in* $\mathcal{G}$ **do**
  $\mathcal{E} \leftarrow$ Get endpoints of group
$\mathcal{D} \leftarrow$ Order and connect $\mathcal{G}$
return D

To define the area bound by the first line of buildings, the approach in algorithm 4.4 is followed. The algorithm requires building data for per side of the river. It computes the visible building nodes from the riverline, and connects these nodes by first selecting visible edges, and then removing hidden edges. Groups of connected visible edges are made, and the start and end points of these groups are found. These endpoints are connected and together with the visible building edges form the first line of building delineation. Figure 4.14 shows the process of the proposed method for building line delineation. The visibility graph determines the visible edges, hidden edges are found (figure 4.14a) and removed, and the remaining visible building edges are connected (figure 4.14b).

4.7. Metrics

Metric values are computed per segment, and table 4.5 shows the list explaining the metrics and the computation of them. General geometric metrics concern the slope and rugosity of the terrain. Slope metrics include ratios of flat, negatively sloped and positively sloped surfaces, and the mean slope value per segment.

For each landuse, the ratio is computed within the segment. Additionally, the landuse diversity and the distance to the first building are computed. The average value for trees is added per segment. The ratio of points with an impervious value > 50 percent over all the points is computed as impervious surface cover. Additionally, the mean value is computed. For flood metrics, the percentage of floodable points is given, the maximum value in the segment, and the mean value. The distance to the last point on the segment categorized as floodable is given as flood reach. A binary value for dike presence is given by intersecting each segment with the dike map. The distance to the first building is added, and if there is no building present, then this value is set to 100 meter. River width is taken as the distance between embankment points. If only one side has an embankment point, then this distance midpoint-embankment point is doubled. Visible space is determined as the ratio of visible points over all points, and visible reach is the distance to the last visible point.

Note that distance to first building, flood reach and visible reach correspond to the three delineation methods described before, but are not computed in the same way. The delineation methods may result in different values than given here. However, since the typology is based on the segments, the computation is changed, and the computation per line segment is meant as an approximation of the delineation values.

One metric is categorical: first landuse. Therefore, this metric can not directly be used in the k-means clustering algorithm. The values for this metric are stored, and can be used to describe variation within a cluster after the clustering algorithm is applied.

Segments that have a water ratio > 50 percent were deemed invalid and discarded. 2546 segments were removed in this step. Flood metrics are chosen as the ratio of impervious surfaces of the segment and the floodable points ratio. Segments with invalid flooddepth values were discarded (flood depth > 30). 26 segments were removed in this step. Segments that did not contain enough elevation data were discarded of. This removed 1137 segments. The final input for the clustering consisted of 148954 segments.

Appendix C shows the distribution of the metric values over all the segments as boxplots, and a table for mean and median values is given. There are a few concerns with these if used as input for k-means clustering

1. High levels of outliers drag centroids of clusters away. This results in metrics with high outliers having a larger influence on cluster results than other metrics.

2. Some metrics have their median at 0 (flood reach, flood ratio, bike, embankment) or 100 (visible reach). This implies that at least half the segments do either not have floodable space, these particular landuses present, or no invisible space.

3. High levels of skewness in many metric distributions. Many metrics are missing a whisker, and have the majority of the observations lay in a particular side of the distribution.

Metric	Description	Computation
Geometry		
Slope	Ratio of section that is flat, that has slope towards river, and that has slope away from river. Additionally, mean slope value.	Compute angle represented by 'rise over run' between points (dy/dx) and work with threshold value (= 0.1 Corresponds to about $\leq 5.71°$) interval for flat surface. Beyond this interval, it is sloped. Compute mean as average angle.
Rugosity	Ratio of the true surface area to the projected flat area	Ratio of length flat area to length curved terrain. Computed for both DTM and DSM elevations.
Landcover		
Landuse ratio	Ratio of each type of landuse over the whole section	Collect all landuses in data. Per type, compute ratio within section.
Landuse diversity	Diversity of land use in section	Number of different landuses in the segment.
First landuse	Landuse of first couple meters next to the river (*categorical*)	Landuse value of first point next to river unless it is water, then it is the next first landuse.
Trees	Presence of trees in area segment lays in	Average percentage of trees in the segment.
Flood		
Imperviousness ratio	Ratio of impervious surfaces	Set threshold for what impervious is since data gives value between 0-100 percent. Take 50 - 50. Then compute ratio.
Imperviousness mean	Mean value of imperviousness on segment	Average imperviousness value of the points.
100 year flood ratio	Ratio of points that are floodable	Floodable points (flood depth value $!=0$) over all points.
100 year flood depth	Depth of flood for floodable points	Maximum for severity and mean value.
Flood delineation	Maximum reach of 100 year flood	Last point on the segment with non-NULL value. If no such point exists, then 0.
Open space		
First building line	Distance to first building	First point with landuse value 'pand'. If no building is present, then this value is set to 100 meter.
River width	Width of river	Distance between embankment points. If only one embankment point is present, then it is taken as twice the value of the distance from midpoint to embankment point.
Visibility	Visible space	Ratio of visible points over all points.
Visible space delineation	Maximum reach of visible space	Last point on the segment tagged as visible. If no such point exists, then 0.

Table 4.5.: Metrics

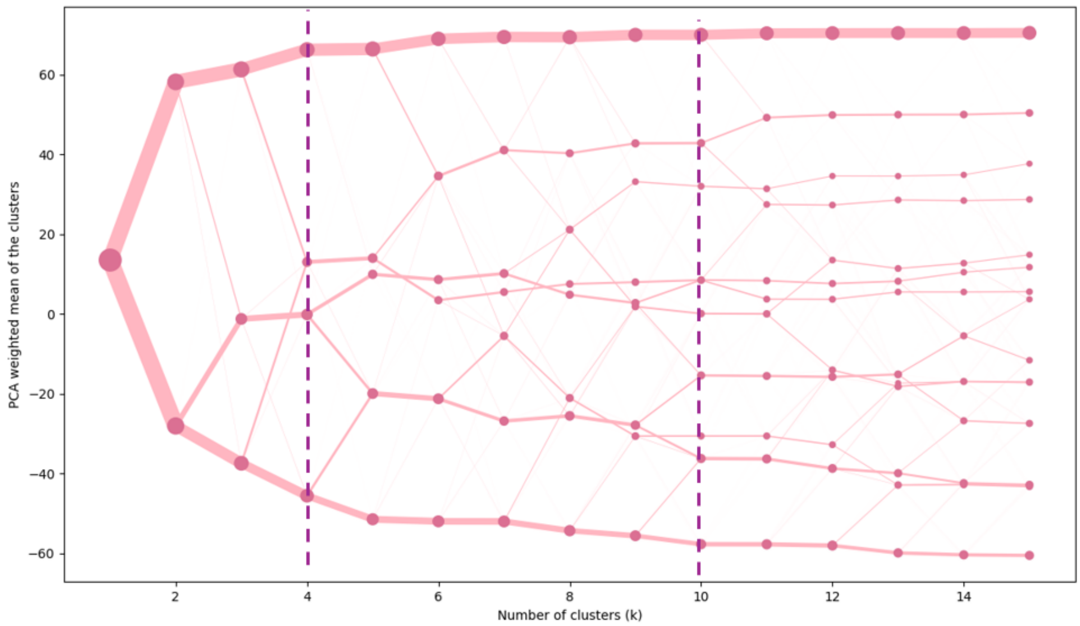
These three problems with the metric distributions can affect the k-means clustering in a negative way, characterizing clusters by outliers. Therefore, we want to make some choices and transform the distributions to be a better fit for k-means clustering.
Firstly, visible reach and flood reach are discarded as input for the clustering algorithm. This is done as these metrics correspond to visible space ratio and floodable space ratio and have their median at maximum and minimum. Is is assumed that those other two metrics provide enough information about the space. These two delineation metrics say more about the possibilities of the extend of the space than about the space itself, and therefore are ruled not meaningful in the typology of river spaces, as the extend of the space is set at 100 meter. However, the distance to the first line of buildings is kept, as this value does not have the median at an extreme.

To deal with outliers and skewness, we want to apply a transformation to the data. As both positively and negatively skewed distributions exists, it is chosen to use the yeo-johnson power transform on all metrics. This transform can be used on positively skewed distributions, negatively skewed distributions, but also normal distributions. Yeo-Johnson normalizes skewed distributions, and transforms data to have mean 0 and standard deviation 1. Appendix C.2 shows the skewness of the metric distributions after transforming them. Yeo-Johnson transform was implemented via scikit-learn *PowerTransformer('method'=yeo-johnson)* on metrics imperviousness ratio, visible space ratio, trees, building, paved surfaces, roads, pedestrian, vegetation, water, yard, landuse diversity, first building distance, flat surfaces, slope, rugosity and river width.

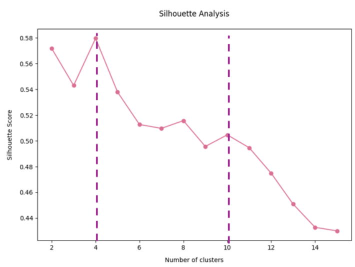
4.8. Clustering

To perform the clustering algorithm, the implementation of the algorithm from ScikitLearn python library was applied using random state parameter 42. This parameter controls the randomness of the initial cluster centroid selection. As k-means is a non-deterministic algorithm, as it randomly initializes cluster centers, it can give slightly different results on different runs of the algorithm. A choice is made to set the random state, to ensure that the same random initialization is done every time the algorithm is ran, to enhance reproducibility of the method.

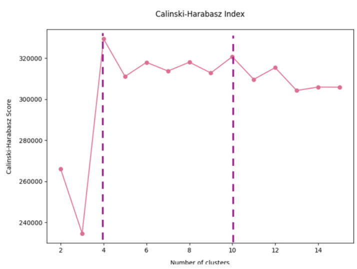
4.8.1. Choosing the amount of clusters



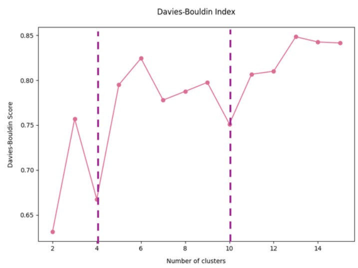
(a) Clustergram



(b) Silhouette score



(c) Calinski-Harabasz score



(d) Davies-Bouldin score

Figure 4.15.: Clustergram and goodness of fit score for k-means clustering for k in $[1, 15]$. Dotted lines are shown at relatively good choices of k ($k = 4$ and $k = 10$)

Figure 4.15a shows the clustergram for the metrics. The clustergram aids the choice of amount of clusters, together with relevant metrics of a goodness of fit such as silhouette score, Calinski harabasz score and Davies bouldin score.
When evaluating the clustergram, we want to look for separation. Did adding an additional cluster bring a meaningful split? The step from 1 to 2 clusters is a clean separation of observations. The split from 2 to 3 already shows some observations switching between clusters, but cluster centers are still nicely separated. This is seen basically in the entire clustergram. $k = 3$ shows one of the clusters splitting into a distinct subgroup, indicating a significant natural separation in the data. Two big clean splits are seen for $k = 4$, and also afterwards clusters seem to become more refined without excessive fragmentation. Observations switch between clusters but the lines are very thin. Between $k = 6$ and $k = 10$ clusters continue to divide, and clusters seem to tangle up a bit. $k = 10$ shows a calmer split, but two potentially meaningful splits are shown in the upper part of the clustergram. After $k = 11$, the bottom part of the clustergram shows some instability, with cluster centers very close together. $k = 14$ seems to provide a relatively clean split, but has small clusters and some centers very close together. Interpreting the clustergram is difficult, as it never shows a truly best option for k, but relatively good choices seem to be $k = 10$ or $k = 14$.
The goodness of fit scores seem to agree on $k = 4$. This value of k gives a maximum for the silhouette score, a maximum for the Calinski-Harabasz score, and a local minimum (and the second lowest value) for the Davies-Bouldin score. A second good option supported by the goodness of fit scores seems to be $k = 10$. The silhouette score has a local maximum here, the Calinski-Harabasz score has its second-highest value here, and the Davies-Bouldin score a local minimum, and the third lowest value.
The evaluation of the clustergram and the goodness-of-fit scores show a great limitation of k-means clustering, as the choice of k is often not so clear. $k = 4$ is better supported by the goodness of fit scores, while $k = 10$ seems to be a bit better looking at the clustergram, where many clusters stay relatively similar. If we are looking for a solution using a more conservative split, $k = 4$ can be used supported by the goodness-of-fit scores. However, this may result in clusters containing many subtypes of river spaces if expected to have a larger range of diversity of river spaces. Therefore, the choice of k for the clustering has to be tied to the expected number of types in the case the algorithm is applied to.

In this thesis, a more detailed version of the clustering is preferred considering this diversity of river spaces in the Netherlands, resulting in a 10-cluster solution, supported primarily by the Davies-Bouldin and Calinski-Harabasz scores, and the clustergram.

5. Results

This chapter describes the results of the delineation methods, the clustering and explains the clustering results as a typology. Previously the segment length was set at a specific value, and this value is evaluated. Similarly for the distance between segments. The metric values distributions are discussed and the clustering results are given and elaborated on. Finally, a typology is developed describing types of river spaces in the Netherlands.

5.1. Segments

Statistic	First building	Visible reach	Flood reach
mean	60.598673	85.771640	27.540937
std	38.198474	26.372051	42.779070
min	0.000000	0.000000	0.000000
25%	21.393035	85.572139	0.000000
50%	64.179104	100.000000	0.000000
75%	100.000000	100.000000	77.611940
max	100.000000	100.000000	100.000000

Table 5.1.: Summary statistics for delineations

The length of the segment was chosen as 100 meter. As a validation check for this value, we can take a look at the defined delineations. The delineations are there to define the boundary of the river space, but were computed using the same buffer of 100 meter. If the segments have the value for this distance often as 100 meter, then 100 meter may not be long enough to ensure covering the whole river space. On the other hand, if the segments often have very low values, then it may be that this length is too long.
Table 5.1 shows the results of the values of the delineations per segment. They describe the distance on the line where the first building was found, what the last visible point was, and the last floodable point. The maximum value was therefore 100, and the minimum 0. One thing to note is that all methods have as maximum 100, which means there are segments where the delineation line probably laid further than 100 meter. If we look at the boxplots in figure 5.1, we can tell this is especially important for visible reach. The median lays at 100, and the 25th percentile lays at 85.57. For flood reach, we see most of the observations have a value of 0. The distance to the first line of buildings has a more spread out distribution. The 100 meter values in this case are often in areas that are mostly vegetation, either in a park in the city or more on the outskirts. The 100 year flood data often either extents way further than 100 meter, or is 0 as many places are considered not floodable.

These results imply that the buffer area of 100 meter can be considered enough for the first-line-of-building delineation. If there are buildings present in the river space, they are

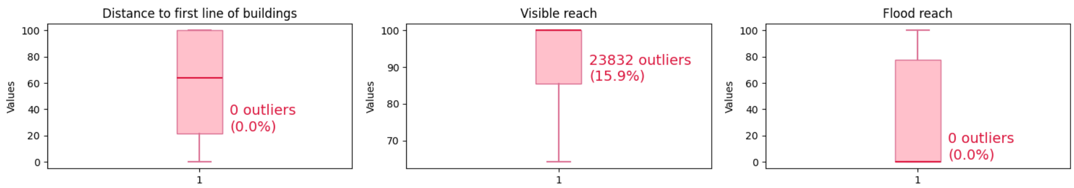


Figure 5.1.: Boxplots for delineation metrics

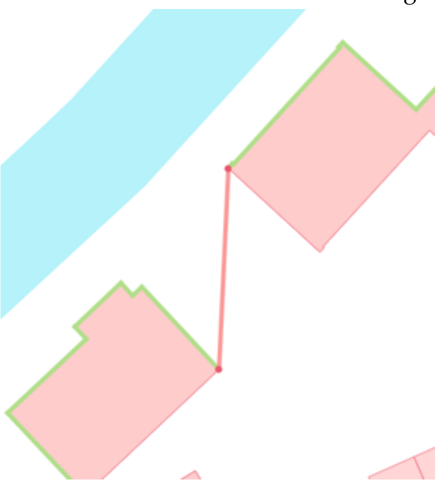
usually within 100 meter distance from the river in the case of the Netherlands. The buffer ensures the delineation provides a space that still has a connection to the river, instead of extending to places where a person could not even see the river. Similar reasoning can be given for flood, as this area can extent far beyond for example the first line of building. For visible space, the buffer could be expanded a bit in most cases. However, in some cases, the large amount of visible space is a result of inaccuracies in embankment height selection. If a building is very close to the river, then the point next to the river may not have a DTM value. This results in selecting a value further away and may be a lot higher than the embankment.

The distance between segments was taken as 25 meter. Figure 5.2 shows some examples on how clusters are distributed over space. Shown is that neighboring segments sometimes have the same types, which is for example visible in the right image, where the segments on the right are all identified as green. However, this is not always the case. The left image shows a large variety between neighbors. Some river spaces extent way past the 25 meter chosen as set value in this thesis, but this value did allow in other cases to differentiate types between neighboring segments.

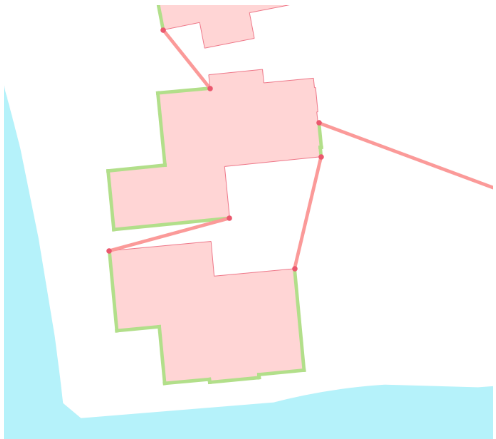
Figure 5.2.: Example of clustering result on the map. Neighboring segments can have the same type, but this is not always the case. Each color is a different cluster.

5.2. Delineation

(a) Example of river space delineation by first line of buildings.



(b) As a consequence of distance between midpoints, the connecting line segment not correctly placed as one building edge is not deemed visible.



(c) Due to the curve of the riverline, the building edge in the middle is deemed visible which may not be desirable.

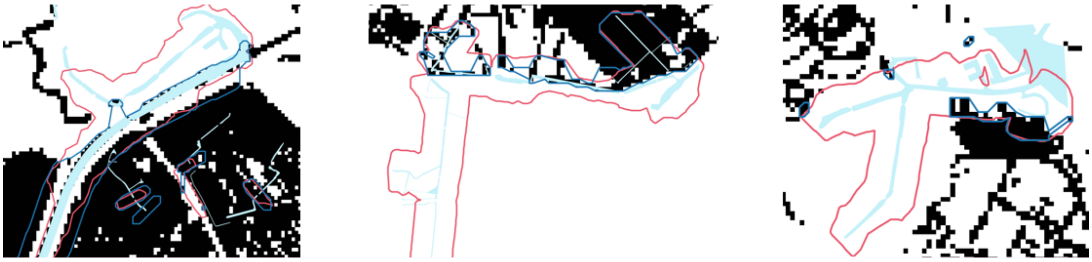
Figure 5.3.: First line of building delineation example. Green edges are visible building edges. Pink edges are connecting edges.

Figure 5.3 shows the delineation algorithm applied to a curved river. Visible building edges are selected (green), and edges are then connected (pink). As the algorithm works with midpoints on the riverline, the placement of these points influence the result. As seen in the second image, it misses a visible building edge. The last figure shows complications occurring at the corner of the curved river. The building edge in the middle of the figure is deemed visible as the right building node can be seen from somewhere far away on the right, but we may not find this edge visible from a qualitative point of view.

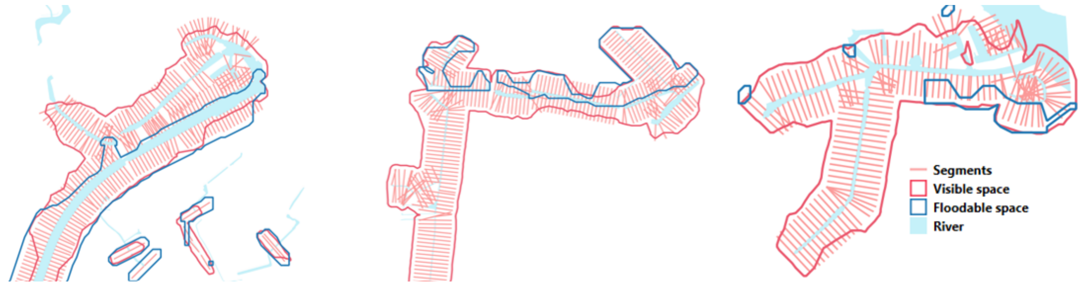
Note that the image on the left shows a connecting line cutting through a river, but this is the result of the method being applied to a single river, and other riverlines are ignored. In order to prevent this from happening, edges that intersect with a riverline have to be removed.

It was chosen to use midpoints on the riverline to reduce computation, as only a few points on the line will be evaluated. However, it may be desirable to determine visibility of the building nodes by projecting the building node to the riverline, and seeing if the this projection line intersects any other edge or building. However, this may cause other issues, including nodes not deemed visible if they have an angled connection to the river. Increasing the amount of midpoints would increase quality of delineation, but also increase computation time.

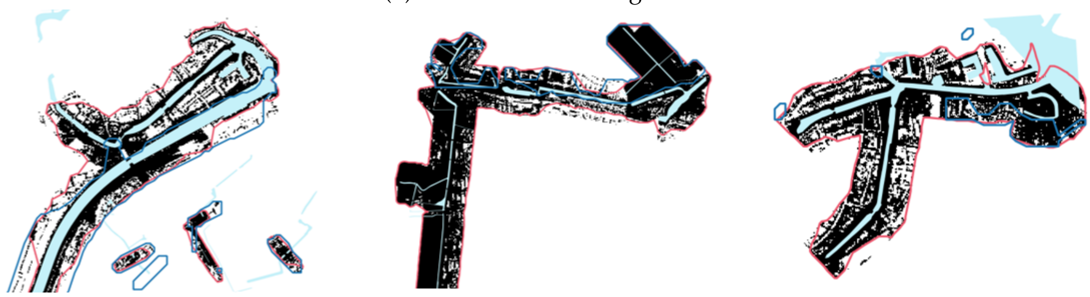
Figure 5.4 shows three examples of the floodable and visible space delineation methods being applied to different rivers in the Netherlands. A value of 1200 was used for α and a minimum area of $625m^2$ was used. The spaces were retrieved by computing an alpha shape around the respective binary raster values. It can be seen that floodable space often extents far beyond the buffer of 100 meter, or no floodable space is present at all. Visible space would often reach the buffer too. As can be seen in the visible space raster images, a lot of smaller visible areas are present, presumably resulting from present trees or similar structures blocking the visibility lines in the viewshed calculations. This results in some invisible space being included in the visible space delineation, especially in the left image. Therefore, the visible space delineation may benefit from a higher value for minimum area.



(a) Flood raster



(b) Delineations and segments



(c) Visible space raster

Figure 5.4.: Examples of result for floodable space and visible space. The black in the raster images shows the floodable or visible points, and white is not-floodable and invisible, or extents past the buffered area.

5.3. Clusters

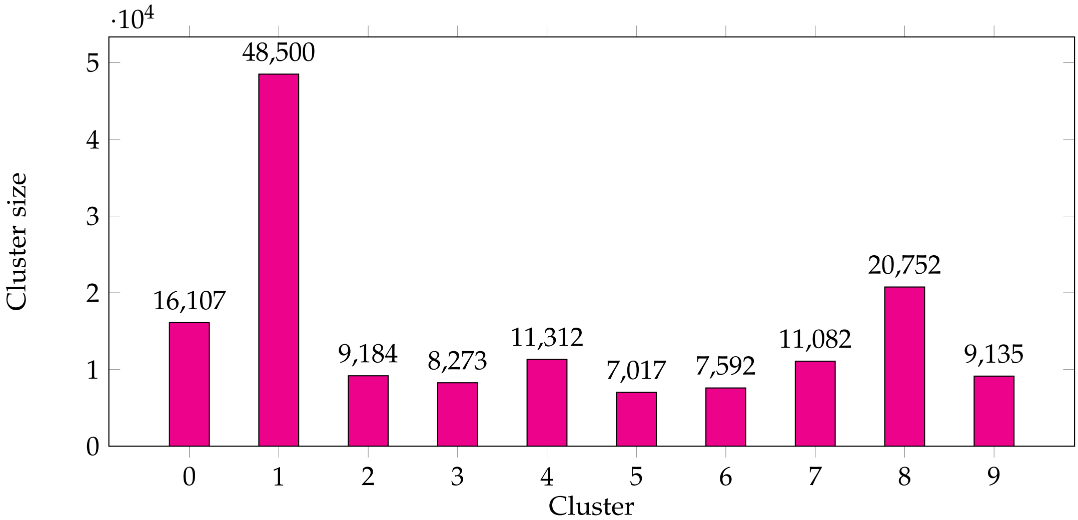


Figure 5.5.: Histogram of cluster Sizes for K=10

K-means clustering was applied to the segments for $k = 10$ using the Yeo-Johnson transformed metrics as explained before, and the ScikitLearn python library implementation of the k-means clustering algorithm. The distribution of the metrics as boxplots for each cluster is shown in appendix D, visualizing the differences and similarities between clusters.
The cluster sizes are shown in figure 5.5. There is one cluster distinctively larger than the others, cluster 1. 5 clusters have less than 10 000 observations in them.

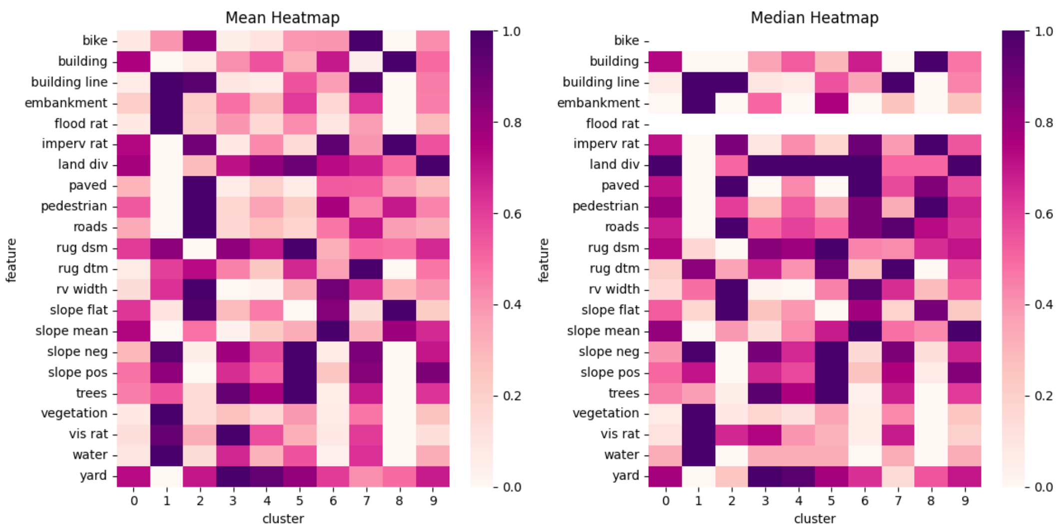


Figure 5.6.: Heatmap of mean and median metric values for each cluster for $k = 10$

The heatmaps of the mean and median values normalized per metric are shown in figure

5.6. This shows what values are distinctive for a cluster compared to the other clusters. For example, cluster 8 has the highest value for metric building. The heatmap, together with the distribution of the metrics per type shown in appendix D and the sampled segments in appendix E, aids the typology development in the next section.

5.4. Typology

To define types of river spaces, each cluster is analyzed. As some clusters show similar characteristics and are hard to differentiate on their own, these are discussed together. Similarities in characteristics are chosen to be based on the building presence and vegetation/tree cover, therefore determined by the balance between the built and the natural environment. This is done as these metrics are especially visible in orthophotos of the space, and can therefore offer meaningful differentiations between main types. Within the type, the differences between the clusters are discussed, analyzing why these are separated by the clustering algorithm and its evaluated in what way they can describe sub-types.

Table 5.2.: Median and mean values per cluster for metrics building, vegetation and trees. Darker coloured cells have higher values.

Cluster	Building		Vegetation		Trees	
	Median	Mean	Median	Mean	Median	Mean
0	31.68	33.94	5.94	8.69	37.13	37.08
1	0	0.38	79.70	69.74	34.65	40.39
2	0	3.25	7.43	13.40	22.28	27.43
3	15.35	18.75	14.36	21.13	56.93	52.53
4	22.77	25.04	9.90	14.19	49.01	47.15
5	12.87	14.71	28.71	29.19	58.42	54.85
6	29.21	31.02	4.95	7.68	21.78	24.65
7	0	2.52	34.16	34.55	46.04	44.73
8	43.07	44.86	0.50	3.30	20.30	22.77
9	20.30	22.31	19.80	20.20	43.56	42.87

Table 5.2 shows the mean and median values for metrics building, vegetation and trees for each cluster. Following this categorization, the following clusters are described together:

- Section 5.4.1 **Urbanized riverbanks**: Contains cluster 0, 6 and 8. High building cover, low level of vegetation, and lower tree percentages.
- Section 5.4.2 **Lush natural floodplains**: Contains cluster 1. Almost no buildings, very high vegetation levels, low to moderate amount of trees.
- Section 5.4.3 **Barren paved land**: Contains cluster 2. Almost no buildings, low vegetation levels, low tree percentage.
- Section 5.4.4 **Residential tree-covered river space**: Contains cluster 3 and 5. Low building cover, moderate vegetation levels and high tree percentages.
- Section 5.4.5 **Balanced riverside corridors**: Contains cluster 4 and 9. Moderate building cover, moderate levels of vegetation, and higher tree percentages.
- Section 5.4.6 **Mixed green strips**: Contains cluster 7. Almost no buildings, moderate to higher levels of vegetation and trees.

The discussion on each type in the following sections make use of visualizations of parameter and metric values. The following plots will be used:

1. Most representative segment as orthophoto and parameter plot. This shows the parameter values for the segment laying closest to the center of the cluster. The legend for the parameter plot is given in figure 5.7 and is consistent throughout the following sections.
2. Spider plot of metrics used in the clustering algorithm. In this plot, metrics are standardized by converting them to z-scores, and then scaled to a suitable range for visualization. The dotted circle represents the average value of the metric over the whole dataset. The dark pink plots the normalized mean of the cluster. The light pink represents the normalized standard deviation within the cluster.
3. Bar chart of first landuse. This plot contains the first landuses next to the river covering 90 percent of the segments in the cluster. The other bins are not shown. Important here is that the bins are more specific than metric landuse categories. In the metric process, some landuse types were merged into one type, such as 'agricultural grassland' and 'grassland' both belonging to more general type 'vegetation'. Bins are coloured by the more general landuse type.

Additionally, appendix D, containing more information on the cluster metric distribution in the form of boxplots and tables, and E, containing nine randomly sampled segments per cluster, will be referred to for describing each type of river space in the Netherlands.

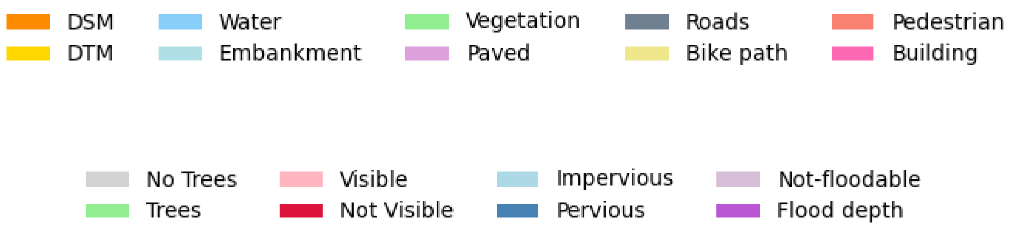
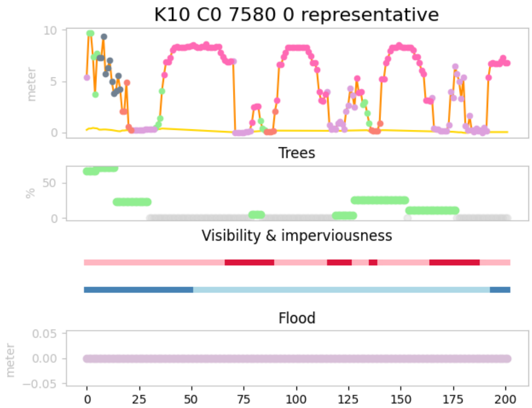


Figure 5.7.: Legend for parameter plot

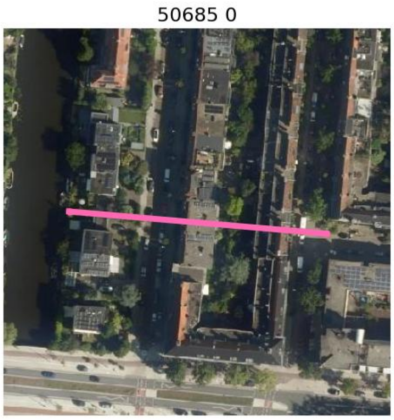
To finalize, appendix F contains the finalization of the findings discussed in this section, providing a reference work for urban river space types in the Netherlands. The 6 types and their subtypes are each summarized on a single page with their average values and most representative segments, offering the typology in an accessible format which can be used as a quick reference when working with urban river spaces in the Netherlands.

5.4.1. Urbanized banks

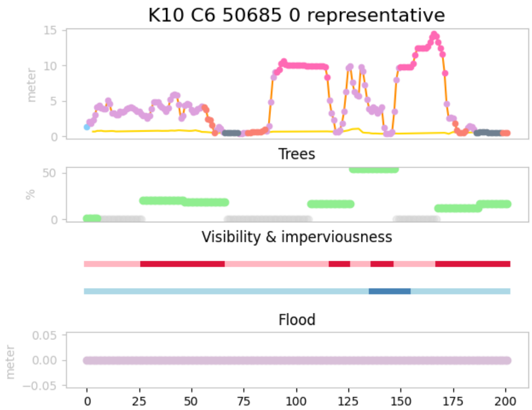
(a) Orthophoto cluster 0



(b) Parameters cluster 0

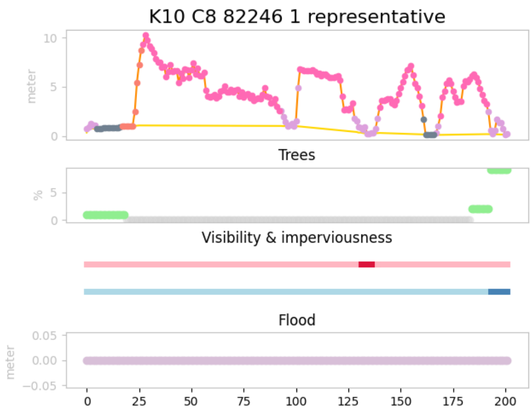


(c) Orthophoto cluster 6



(d) Parameters cluster 6

(e) Orthophoto cluster 8



(f) Parameters cluster 8

Figure 5.8.: Most representative segments of urbanized river banks.

The first identified type of river spaces in the Netherlands encompasses densely built residential river spaces, resulting from cluster 0, cluster 6 and cluster 8. These spaces have a high building cover, and low vegetation and tree percentages compared to the other clusters. They are densely built, residential or industrial areas. The clusters in this type have elevated or moderate levels of paved space, pedestrian paths and roads. Therefore, the level of impervious surfaces is also higher than in other types. the representative segments in figure 5.8 show the space almost fully covered in impervious surfaces (light blue), but the most representative segment for cluster 0 does show some pervious green space right next to the river. The space tends to not be floodable, with every representative segment following this trend.
The terrain in this space is mostly flat. Negative and positive slope percentages are lower than in other clusters. The rugosity values are also similar and are leaning more to lower values.

The clusters also present differences in metric values. Cluster 8 has a higher level of building cover, which is already visible in the most representative segments in figure 5.8. The lower building cover in cluster 6 is also visible, with the first line of buildings also being further away from the water in this cluster. The limited amount of trees and especially vegetation in cluster 8 can also be seen, although the segment for cluster 6 also does not contain vegetated areas.

The sampled segments for cluster 0 in appendix E show residential neighborhoods, covered in buildings. Some green spaces are present, especially right next to the river, and the space is often partly covered by paved or pedestrian areas. The samples for cluster 6 show more of a mix between residential and industrial areas, with less green spaces present than in cluster 0. The building cover is slightly lower, with the first line of buildings being pushed back in these spaces. Very limited green space is visible, and if present, it presents itself mostly right next to the river. The samples for cluster 8 show mostly industrial areas covered in buildings. Almost no vegetation or trees are present. The spider plots in figure 5.11 visualize the similarities between the clusters quite well, often peaking at similar spots. The differences are visible in the peaks at the flat and mean slope, the slight differences in pedestrian, paved and road values, and tree peak, which is for all under the average value, but significantly less for cluster 6.

The first landuse bar charts show that often some green space is present right next to the river for cluster 0 and 6, but often some (potentially) paved areas (yard, pedestrian path, open pavement, parking area) is located right next to the river. The first landuse bar chart looks very similar for cluster 6 as the one for cluster 0. The bar chart for cluster 8 shows that the space next to the river is most often covered in some sort of paved or pedestrian area. Green spaces and ditch banks exist, but less often than in cluster 0 and 6. This space also sometimes has a building directly next to the river, which is not seen in cluster 0 and 6.

Following the discussion above about metric values, sampled segments and first landuses, the three sub-types of cluster 0, 6 and 8 in urbanized banks can be described as follows:

- **Urbanized riverbanks: Mixed hardscape greenway** A building-covered residential river space with low vegetation levels but with more trees present. Green space is primarily present right next to the river
- **Urbanized riverbanks: Open-edge urban river space** With buildings set back, this building-covered river space is more open towards the river, which allows bike paths

(a) Cluster 0 (b) Cluster 0

(c) Cluster 6 (d) Cluster 6

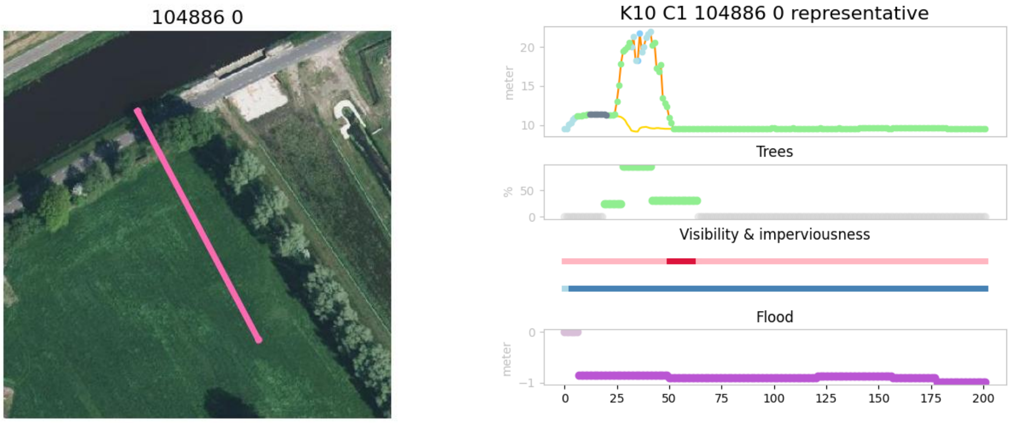
(e) Cluster 8 (f) Cluster 8

Figure 5.9.: Spider plots and first landuse bar charts for urbanized banks.

to often be present in this space. Vegetation and tree presence is minimal, and the space tends to contain more open paved surfaces.

- **Urbanized riverbanks: Dense built-up river space** A space almost completely covered in buildings, limiting the visible space, and offering very limited room for vegetation and trees.

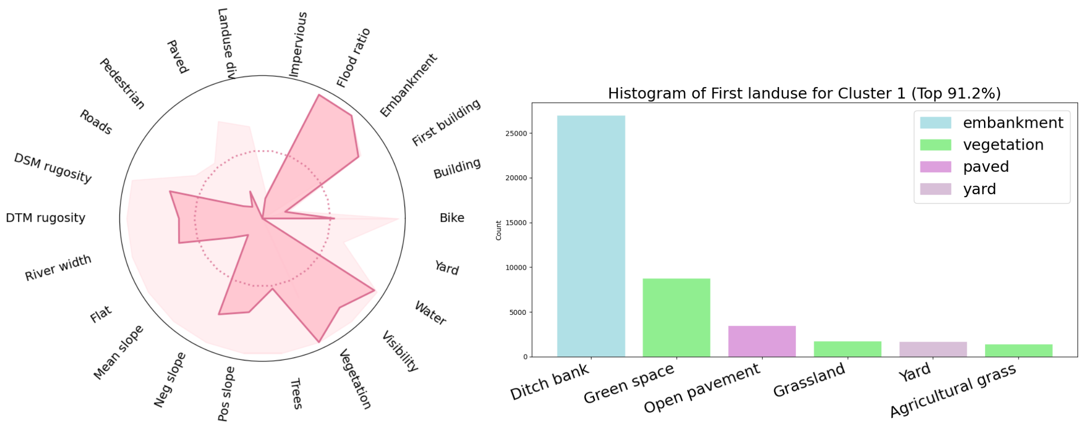
5.4.2. Lush natural floodplains



(a) cluster 1 (b) cluster 1

Figure 5.10.: Most representative segment of green terrain

This type based on cluster 1 is a green area with basically no buildings present. The space has a high chance of being floodable, and contains very little paved surfaces, pedestrian paths or roads. The terrain is often sloped and the rugosity indexes are relatively high. The space has high levels of vegetation and a high tree presence. The most representative segment in figure 5.10 shows an area covered in grass with some tree-lines. Similar situations are seen in the sample in appendix E, which show green, non-built areas with possibly smaller rivers present next to the main river. The exception is the sample with ID 102882 side 0, which is mostly covered in paved surface. Some buildings are visible, but the space is still very green. This river space has a lower value for landuse diversity, which is a result of the exceptional high vegetation cover, visible in both the most representative segment and the sampled segments.



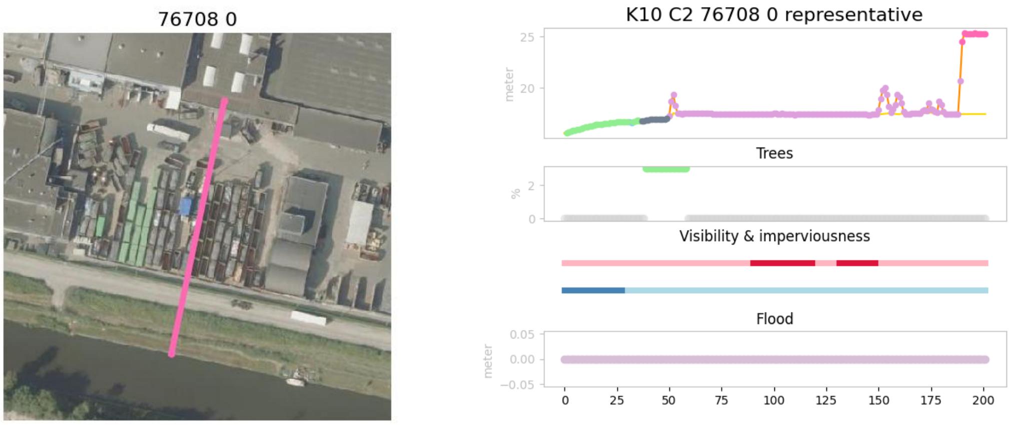
(a) Cluster 1 (b) Cluster 1

Figure 5.11.: Spiderplot and first landuse bar chart of green space

The spider plot shows most metric values being either far above average or far below average. Only trees and bike paths have values closer to the mean. The first landuse bar chart shows very green areas right next to the river, and it almost never a more paved landuse type. This river space type in the Netherlands can be summarized as follows:

- **Lush natural floodplain** A green, often floodable, river space absent of buildings or has limited urbanized areas present. The green area is mostly covered by low vegetation instead of trees, and therefore the river is often visible wherever you are in the space.

5.4.3. Barren paved land

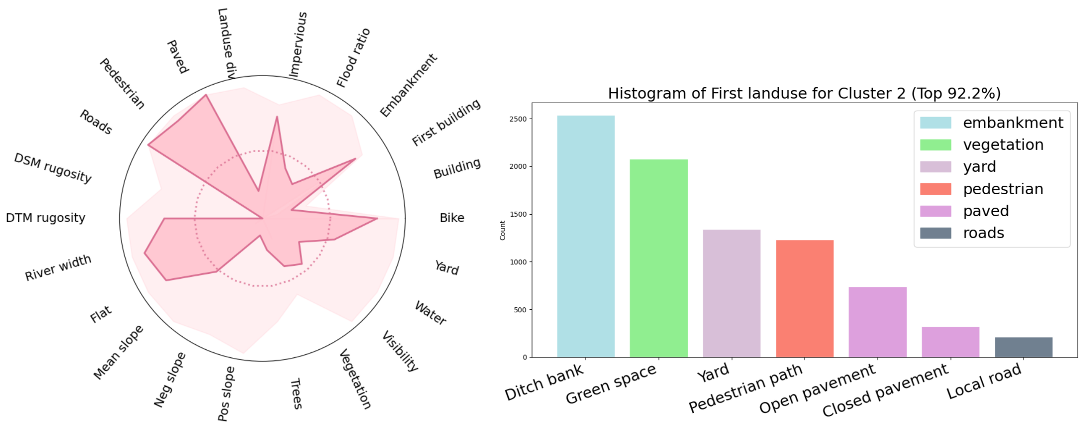


(a) Orthophoto cluster 2 (b) Parameters cluster 2

Figure 5.12.: Most representative segment of barren paved land

This type has a very low building cover, and if there is a building present in a segment, it is far away from the river bank. It distinguishes itself by having relatively high values for road presence, pedestrian areas, bike paths and paved surfaces, thus also a higher level of impervious surfaces. Trees and vegetation can be present, but in relatively low quantities. The type has a relative high DTM rugosity value, but low DSM rugosity.

One note to make about the high levels for metrics road or bike, is that sometimes it is the result of a limitation of cross-section segmentation. For example, the section can align with a road, and therefore gives a false high value for road in the area. This is visible in sample with id 50610, side 0, appendix E. The other samples in the appendix do show more accurate representations of the space, as presented are grey areas with limited building cover and vegetation, with high levels of paved and pedestrian surfaces.



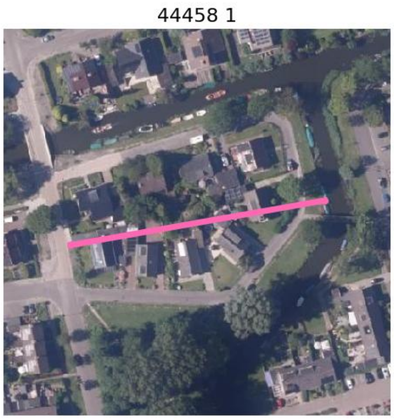
(a) Cluster 2 (b) Cluster 2

Figure 5.13.: Spider plot and first landuse bar chart for barren paved land

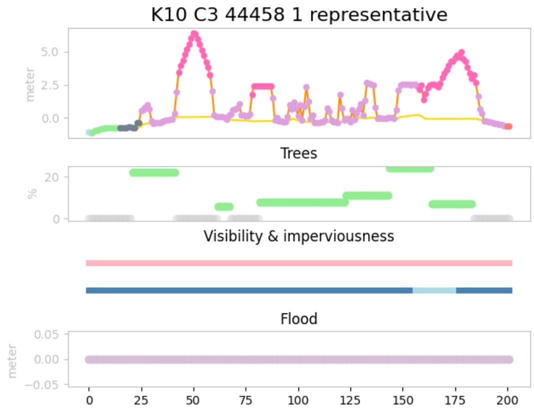
The first landuse bar chart in figure 5.13 shows that a green space is often present next to the river, but there are also a lot of cases where the first landuse is some sort of paved area or even a road.
This river space type can be summarized as follows:

- **Barren paved land** A grey, flat river space, with limited vegetation and buildings. Space is often covered by pavement, roads, pedestrian areas, yard or bike paths and therefore has a high level of impervious surfaces.

5.4.4. Residential tree-covered river space

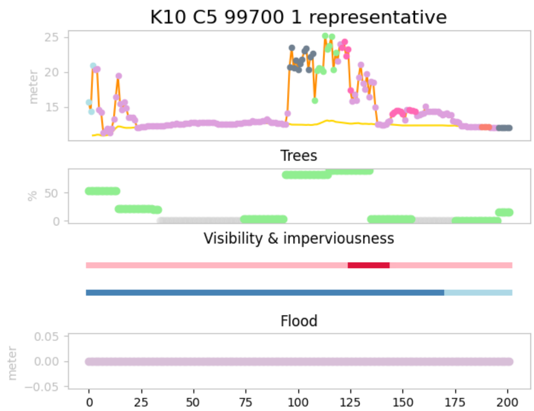


(a) Orthophoto cluster 3



(b) Parameters cluster 3

(c) Orthophoto cluster 5



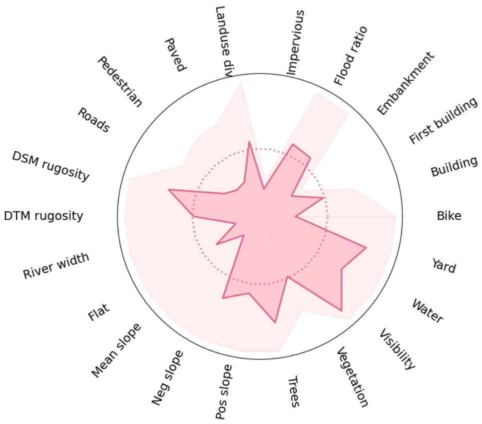
(d) Parameters cluster 5

Figure 5.14.: Most representative segments of residential tree-covered river spaces

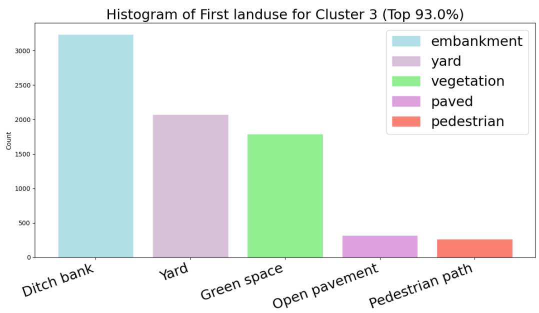
Residential tree-covered river spaces have lots of trees present, and some vegetation. The space has some built areas, but a lot less than the type of river space described in section 5.4.1. This type differentiates itself from type 'urbanized banks' by this lower building density and in general has higher tree and vegetation levels.

This type resulting from cluster 3 and 5 contains a small amount of buildings, and have a high level of trees present in the area. Vegetation can be present but is not a defining factor, and both representative segments in figure 5.12 show almost no vegetation. The photos do show the high amount of trees present. Buildings are present, but the area shows a balance between the built environment and vegetated areas. The space can be partly floodable, which the representative segments do not have, and has very low impervious surface cover, which is indeed shown with very little dark blue present and the bar being mostly light blue. The space does not have to be completely flat, but can have some sloped surfaces.

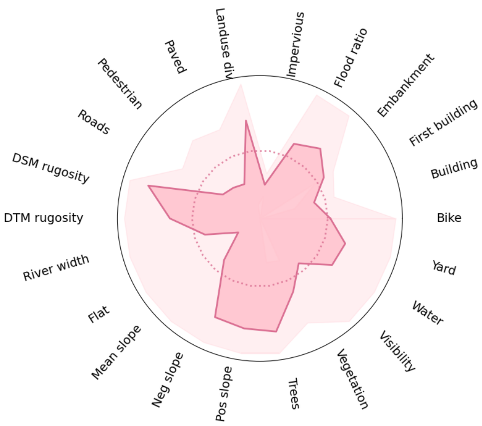
Cluster 5 has a higher level of vegetation, accompanied by a lower building cover than cluster 3. Thus in general, cluster 5 shows more vegetated areas than cluster 3. However, these two sub-types are very similar. This can especially be seen in the samples in appendix E, with most photos showing green, partly urbanized areas, with often many trees present. A difference than is quite visible in the samples is the sloped areas of cluster 5, with often some kind of bump near the river. This is absent in the samples of cluster 3. Cluster 3 shows its buildings closer to the river than the samples in cluster 5.



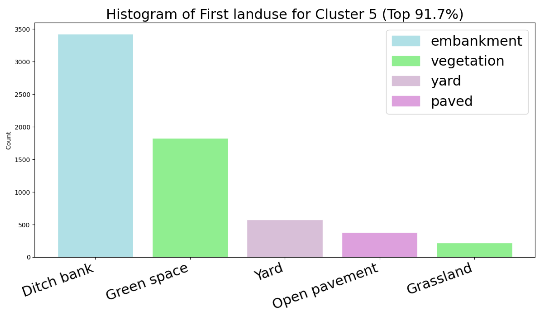
(a) Cluster 3



(b) Cluster 3



(c) Cluster 5



(d) Cluster 5

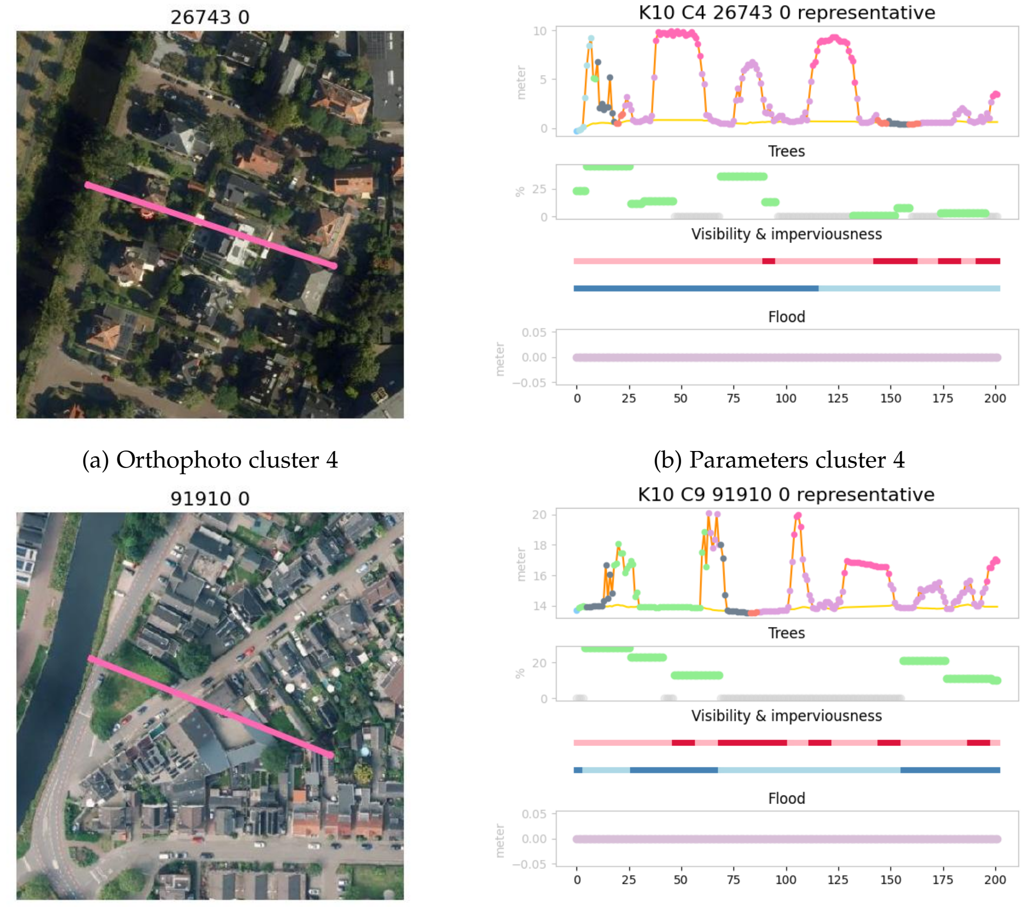
Figure 5.15.: Spider plots and first landuse histograms of residential tree-covered river spaces

The spider plots in figure 5.15 visualize the similarities between cluster 3 and 5 quite well. Similar peaks and lows are seen. The main difference in sloped surfaces is also visible, and peaks in the spiderplot for cluster 5 seem to be more prominent than in cluster 3, deviating away from the average values over all observations. Cluster 3 shows the peak for visible space ratio, which is absent in cluster 5. The first landuse bar charts for both clusters are similar, with cluster 3 having more yard (space associated with a building) present next to the river. This cluster can also have a pedestrian path present near the river.

Following the discussion above on metric values, sampled segments and first landuse distributions, the two sub-types defined by cluster 3 and 5 can be described as follows:

- **Residential tree-covered river spaces: visible areas** A tree covered river space with lower building cover, consisting of a mix of flat and sloped terrain. The space consists mostly of pervious surfaces. Most of the area in this type is considered visible space.
- **Residential tree-covered river spaces: sloped areas** A tree covered river space with low building cover, characterized by the presence of sloped areas. The space consists mostly of pervious surfaces.

5.4.5. Balanced riverside corridors



(a) Orthophoto cluster 4

(b) Parameters cluster 4

(c) Orthophoto cluster 9

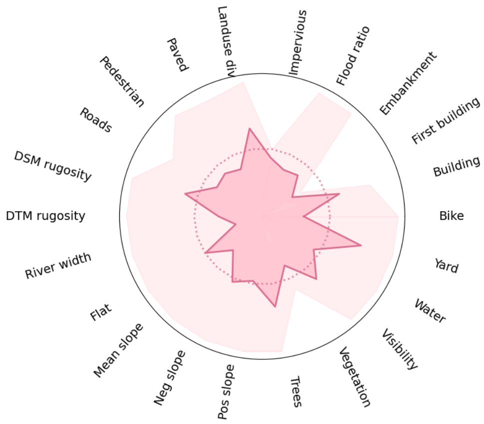
(d) Parameters cluster 9

Figure 5.16.: Most representative segments

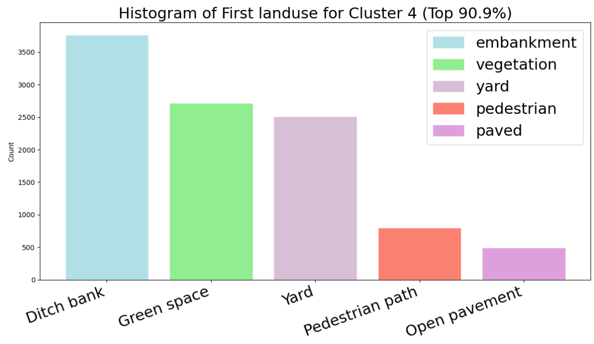
The balanced riverside corridor is characterized by a moderate building presence, with relatively low levels of vegetation, but higher tree percentages. Impervious surface ratio is quite high, thus paved surfaces, pedestrian paths, yard, and roads can be present, and these values tend to be higher in cluster 9 than in 4, except for landuse type yard. Slope and rugosity values are similar for both clusters, with cluster 4 having a slightly lower value for DTM rugosity. Most metric values are very close to average metric values over all observations.

Cluster 4 describes a more green type of river space compared to cluster 9, which is especially visible in the most representative segments in figure 5.16. Both show residential areas, but the area of cluster 4 has higher tree coverage and less road or paved area coverage. The spiderplots in figure 5.17 show very average values for most metrics, especially for cluster 9. A high value for both clusters is seen at landuse diversity. Cluster 9 has a bit more sloped surfaces and thus also a higher value for DTM rugosity.

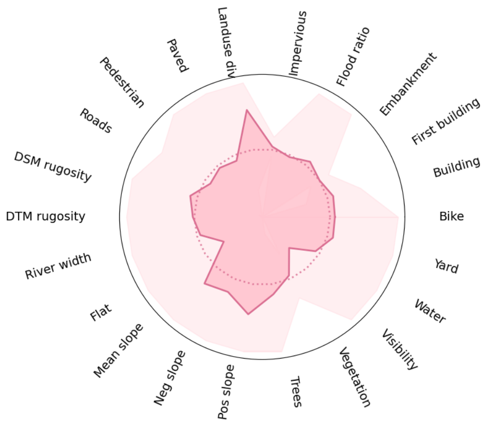
The first landuse bar charts for both clusters are very similar. Cluster 4 has more 'yard' next



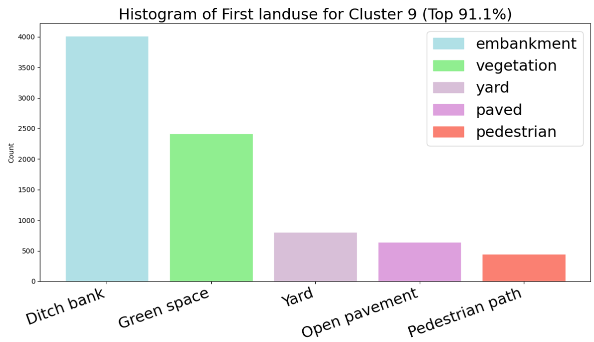
(a) Cluster 4



(b) Cluster 4



(c) Cluster 9



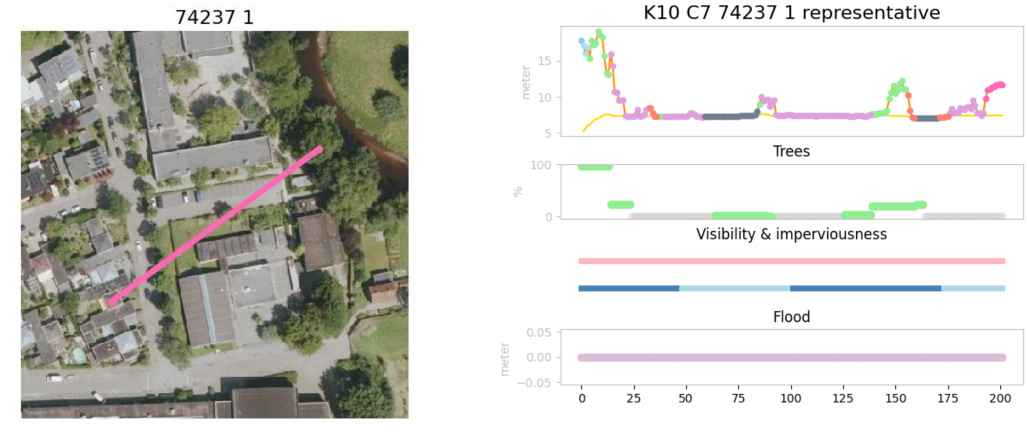
(d) Cluster 9

Figure 5.17.: Spiderplots and first-landuse bar charts of moderate-density residential mixed space

to the river and higher possibility of there being a pedestrian path. In general, cluster 4 has elevated 'yard' ratios, which is also prevalent in the sampled segments in appendix E.

- **Balanced riverside corridors: Buildings and yards** A river space containing a moderate amount of buildings and their associated yards, often near the river.
- **Balanced riverside corridors: Set-back buildings and green riverbanks** A space with a moderate amount of buildings, positioned further away from the river, often allowing a green area next to the river covered by vegetation or trees.

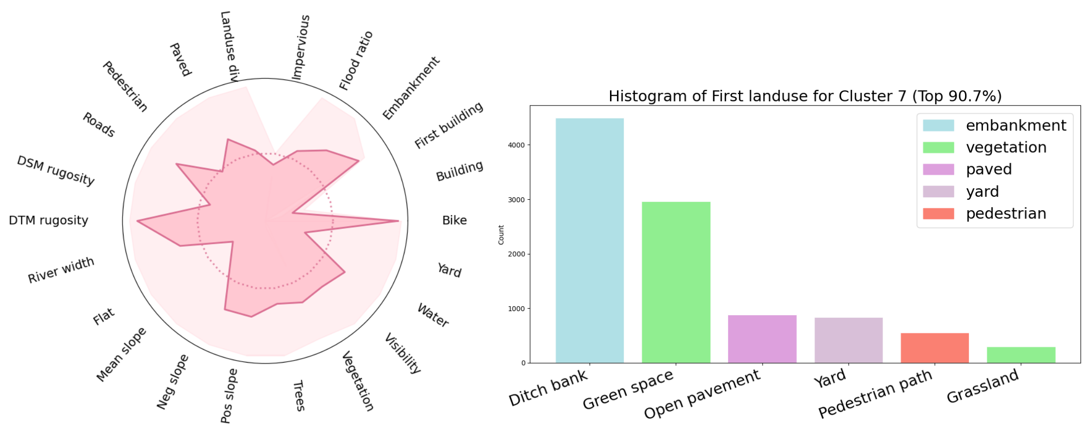
5.4.6. Mixed green strips



(a) Orthophoto cluster 7 (b) Parameters cluster 7

Figure 5.18.: Most representative segment of mixed green strips

This type coming from cluster 7 defines itself by no (or very low) building presence in the segment. There is green space present, but not overwhelmingly as we saw in type 'lush natural floodplains' described in section 5.4.2. Samples show areas where buildings are often present, but the segment does not cross them (such that the segments lays right in between some buildings, or its an open area surrounded by buildings), and lots of trees or vegetated areas are seen, but also much paved area. The metric distributions show larger values for paved areas and roads. This is also visible in the sampled segments. Many of them have large pieces covered in paved space. The orthophotos often show present roads. This cluster has the highest average value for DTM rugosity and for bike paths, which is not as visible in the samples.



(a) Cluster 7 (b) Cluster 7

Figure 5.19.: Spiderplot and first-landuse bar chart of mixed green strips

The spiderplot shows these strong peaks for DTM rugosity and bike paths. It also shows the

balance between vegetation and trees, compared to paved, roads and bike, as a mixed type. All these metrics have values higher than average. The first landuse bar chart shows lots of green space right next to the river, but also more paved areas being present.
This river space can be summarized as follows:

- **Mixed green strips** A vegetated river space laying in the built environment, often partly sloped, characterized by the extremely limited amount of buildings in the section, and the balance between vegetated and paved areas.

5.5. Typology validation

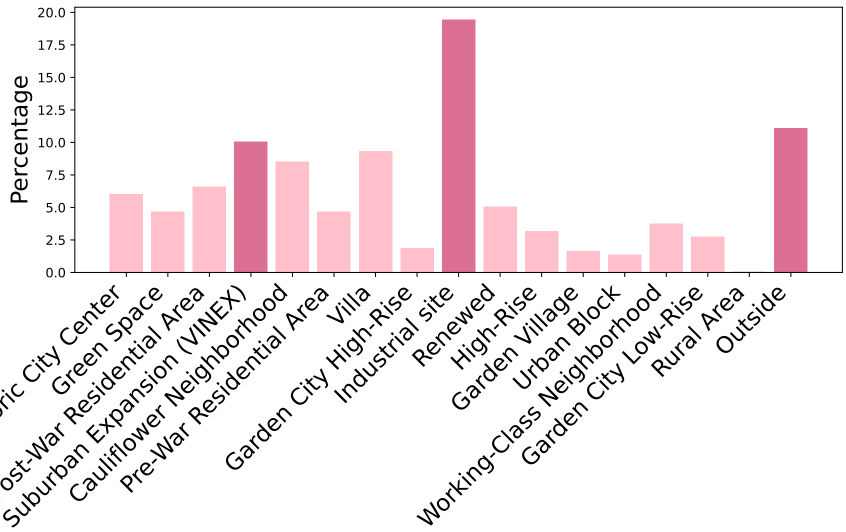


Figure 5.20.: Distribution of all segments for neighbourhood type

The neighbourhood typology (*wijktypologie*) created by the 'Hogeschool of Amsterdam' can give further insights in the quality of the developed typology, and provide extra information on the relationship between river space type and location. This typology classifies urban neighborhoods based on characteristics that are tied to the time period the neighborhood was build in, and can be used to determine how a neighborhood deals with extreme weather and floods. The ratios for how many of the segments per cluster fall into a specific type are computed and shown in the following histograms. This step is done moreso to give an extra dimension to the river space typology.

The histogram plots for each cluster are shown below, describing how much of the segments fall into each type of neighborhood. The darker bins identify categories for which at least 10 percent falls in that category. One thing to note is that all clusters have at least 10 percent lay in Industrial sites. When looking at the distribution of all segments combined in figure 5.20, we see that the largest category over the whole dataset is also Industrial site. This also shows a large category called 'outside'. This column means it falls outside of the bounds of the typology, such that it is unclassified. These are often the outskirt vegetated urban areas.

Cluster 0 has peaks at Historic city center, VINEX neighborhoods, Cauliflower neighborhoods and Industrial sites. Comparing this to cluster 6, which has its only real peak at Industrial sites. Cluster 0 seems to have their distribution more spread out over the categories, while cluster 6 has almost a third of its observations fall into Industrial sites. For cluster 8, we can see a large peak at 'Historic city center', with about 17 percent. Over 20 percent lays in Industrial sites. This is the only type that has a strong peak at this category.

Cluster 1 has a very strong peak of 40 percent of its observations laying outside the neighborhood typology. Besides not having a neighborhood type, observation often lay in green

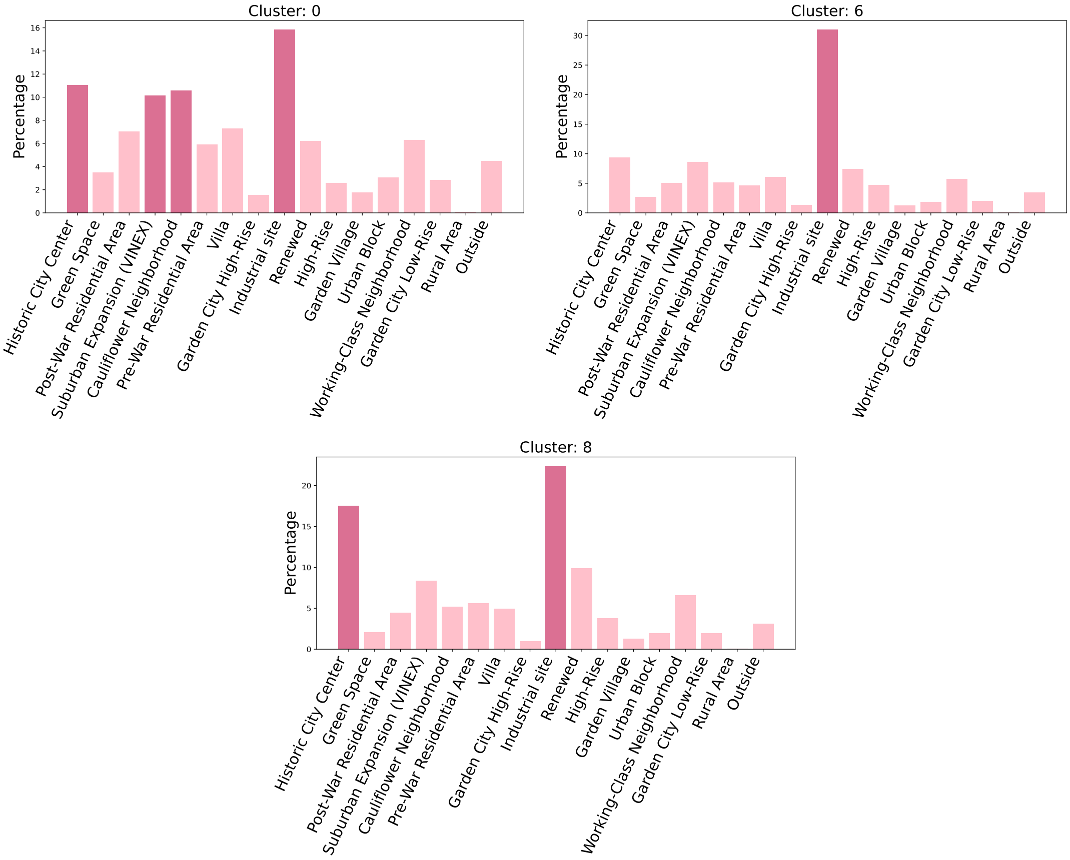


Figure 5.21.: Histograms per cluster for neighborhood typology for cluster 0, 6 and 8

spaces, Villa neighborhoods and Industrial sites. Cluster 2 has a very strong peak of over 35 percent at Industrial site. The rest of the categories often lay below 5 percent.

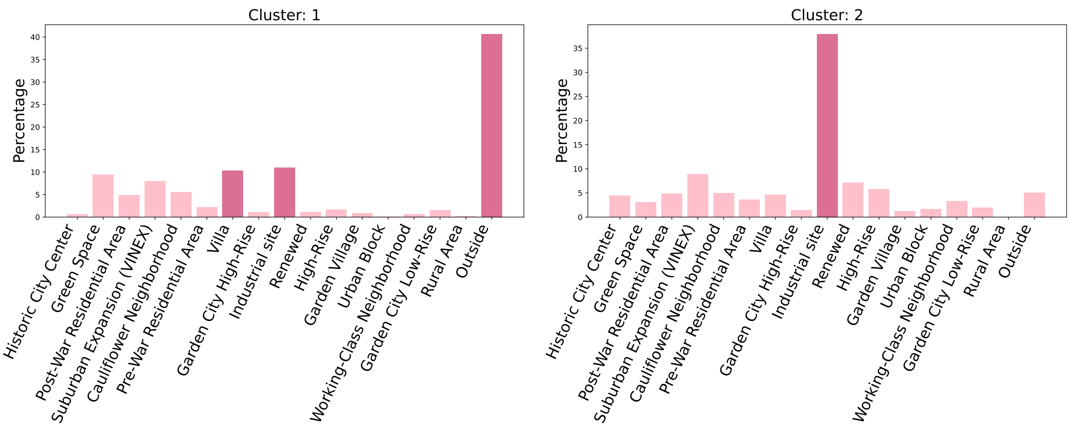


Figure 5.22.: Histograms of neighborhood typology for cluster 1 and 2

Cluster 3 and 5 show quite similar histograms. Both have at least 10 percent of observations lay in VINEX, Villa, Industrial sites or outside the typology. Cluster 5 is slightly more spread out with also having about 13 percent of observations lay in Cauliflower neighborhoods, while cluster 3 has 9 percent laying in this category.

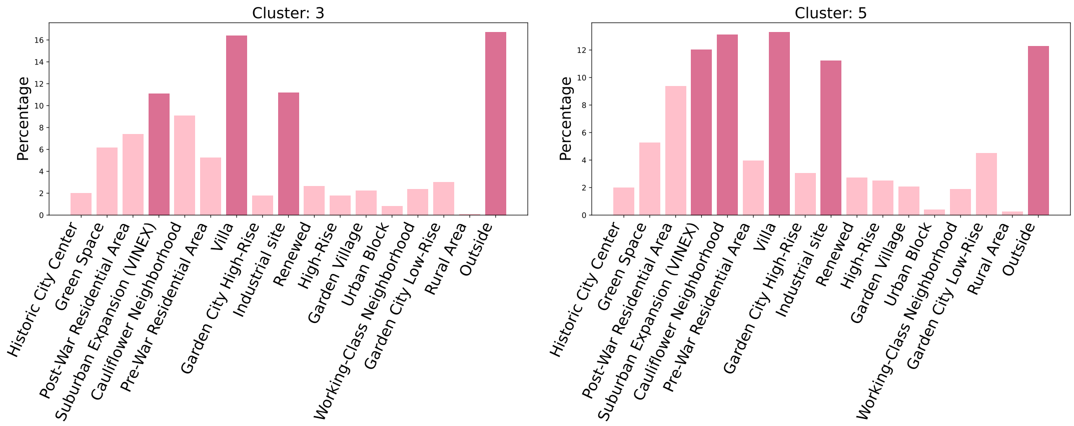


Figure 5.23.: Histograms per cluster for neighborhood typology for cluster 3 and 5

Clusters 4 and 9 in figure 5.24 again show similar peaks, this time at VINEX, Cauliflower neighborhoods, and Industrial sites. Cluster 4 also has 12 percent of its observations lay in Villa neighborhoods, and has some more spread over the types.

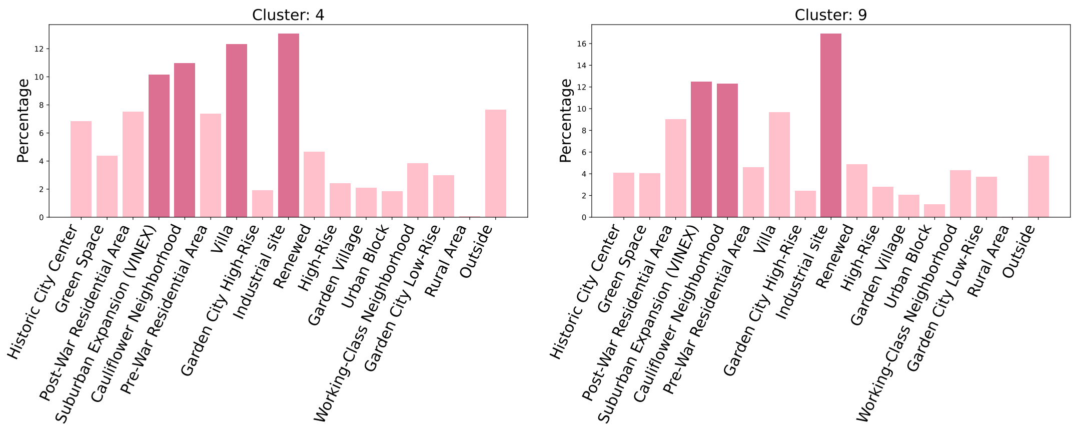


Figure 5.24.: Histograms per cluster for neighborhood typology for cluster 4 and 9

Cluster 7 has over 20 percent of its observations lay in Industrial sites. Larger peaks are also seen outside the typology, for VINEX neighborhoods, Cauliflower neighborhoods and Villa neighborhoods. Every histogram per cluster has different characteristics, but similar clusters tend to show similar histogram distributions, such as 3 and 5 or 4 and 9. Some clusters have their observations more spread out, such as cluster 0, and others show one main peak, such as cluster 1. The differences in distribution between clusters show that the river space typology corresponds to differences in neighborhood typology.

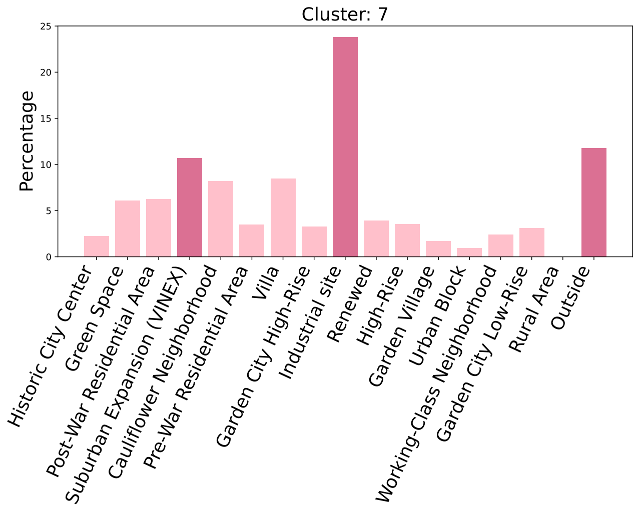


Figure 5.25.: Histogram of neighborhood typology for cluster 7

6. Discussion

The representation of the urban river space remains a complex and evolving spatial challenge. The research conducted in this thesis contributes to the field by exploring how these spaces can be delineated and understood. The three delineation methods – the first line of buildings based on visible building nodes, the visible space as an alpha shape around a viewshed, and the floodable space as an alpha shape around the 100 year flooddepth grid – show different perspectives on how the space can be understood and assessed. Each method reveals different spatial dynamics, highlighting the multi-dimensional nature of river spaces. Furthermore, by developing a data-driven urban river space typology based on cross-section analysis, this thesis presents a new way to categorize these spaces, which can be used by planners and designers to interpret and understand the variety of urban river spaces in the Netherlands.

The thesis focuses on the Netherlands, and therefore uses datasets available exclusively in this country. Scaling the methodology to a broader context would require relying more on universally available datasets, such as OSM. The method may also have to be changed if certain datasets are not available at all in a region. The Netherlands offers high resolution data, for example the DSM with a resolution of 0.5 meter. This has benefits, as it is more accurate than lower resolution data, but this may also be necessary computationally expensive. Lower resolution data is coarse, but may miss important information. This thesis ended up using a 5 meter resolution DSM to perform the visibility analysis, as the 0.5 meter DSM took a very long time to run all urban rivers in the Netherlands. Small tests showed not a very big difference between the two viewsheds.
As mentioned in the implementation chapter, many different datasets with different resolutions and qualities were used. This can cause problems, and did, which were taken care of during development. It must be noted that the method is highly dependent on the data used, and therefore data quality problems affect results. Misalignment and differences in resolutions were seen, and implementations of the method should be aware of this.

The developed delineation methods were applied to urban rivers in the Netherlands. The visible space delineation was performed on a 5 meter resolution raster resulting from viewshed computations within a 110 meter buffer. It was found that the visible space often reached this buffer limit. 110 meter was chosen in this thesis as this is a distance that is expected to ensure a visual connection on a clear day. The buffer value is a threshold that needs to be chosen by the user of the method, and may depend on country or area the delineation is performed in, and may be subject to future research.
The floodable space delineation based on a 25 meter 100 year flood depth data within a 100 meter buffer around the river revealed that segments often are either completely floodable or not floodable at all. River space delineation based on floodable space may therefore benefit from a more nuanced determination of floodable space allowing more variation, or can be used in combination of other delineation methods.
Visible and floodable space were both computed from a raster using an alpha shape, setting thresholds for minimum area and for α. These thresholds may be subject to future work,

optimizing the method to work for other areas.
An algorithm for first line of building delineation based on visible building nodes was described and applied to areas in the Netherlands. Shortcomings of the method were described in the results chapter. The method may have to be refined to accommodate for edge cases, to provide solutions for described limitations, and to scale-up the method to be applied to full cities instead of singular riverlines.

The delineation methods were not used in the typology, but typology and delineation were developed separate from each other. This was a result of the structure of the process of writing the thesis. Ideally, segments would have been cut at the delineation bounds, to develop a typology of urban river spaces in the Netherlands defined by first-line-of-buildings, visible space or floodable space. The typology offered in this thesis includes information beyond these boundaries (or possibly does not extent towards this boundary). However, the same method as proposed in this thesis, combined with a delineation method cutting the segments, can be used to develop such a typology.

This thesis investigates the river space as the area in direct contact with the river. The adjacent areas are not considered, even though they may have a large influence on the space and on the whole river corridor. In the existing 2D definition of the urban river space by Forgaci [2018], it is only a part of a bigger river corridor description. Opportunities for further research lay here in terms of detailing other parts of the urban river corridor delineation, making use of available 3D data and extending the definition of the space. Expanding the scope to include areas beside the river space would provide a more comprehensive typological framework.

The typology relies on specific parameter and metric choices, to which there is an inherent subjectivity. Although parameters were carefully considered and selected based on literature review and the objectives of this study, there is still a certain degree of uncertainty. Future work could explore the influence of using different parameters and metric values, and possibilities for expanding the method. As the goal here was to aim for a base structure of the river space based on spatial characteristics, a lot more information can be added to gain better understanding of the behavior of the space.

This thesis uses cross-sections as segments representing the type of river space, which has its limitations. Even though this thesis took an effort in retrieving 'clean' segments, this was not always the case. There were segments in the dataset that laid more next to the river than perpendicular to the river. This is a result of branched rivers. Future research would have to account for the limitations of the used approach, and may include improving the segmentation method. While cross-sections provide valuable lateral information such as slope and elevation differences, they may not be very effective in capturing longitudinal continuity along the corridor. Other methods of segmentation would have to be researched to draw conclusions on the effectiveness of the cross-section segmentation. Additionally, a cross-section may lay, for example, in between buildings. Even if the space around the segment is covered in buildings, the segment will output no building presence. In this case, the cross-section segment does not represent the space well.
It must be noted that the creation of the typology based on lateral cross-sections does not directly take into account the longitudinal direction of the river space. However, information on this can be retrieved by analyzing the different types along the corridor, including diversity, fragmentation, or which types are most present. This is how the proposed typology created information for the entire corridor, not only in one direction. This analysis was not performed in this thesis, but may offer opportunities in further research. Figure 6.1 visualized this process. A pattern of cross-section segment types can identify a corridor river

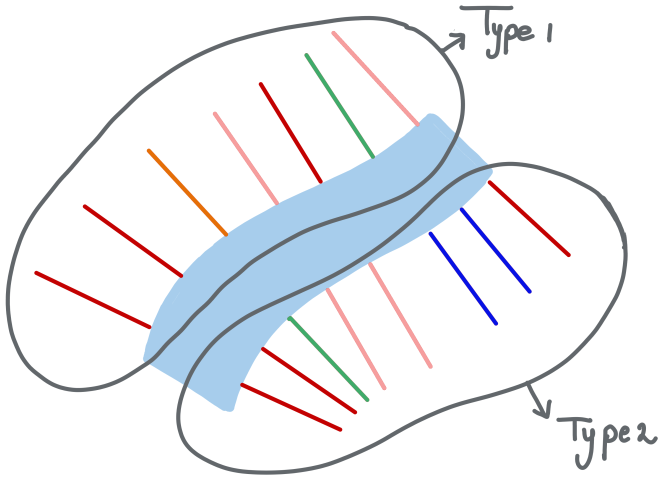


Figure 6.1.: Types in this thesis are defined per section (the colored lines). Types of corridors can be determined by combining them.

space type. This can also solve the issue of the segment not representing the space well in cases where it falls in between buildings, as these segments are included in the type and not a type on its own. Another option to reduce the impact of this problem would then be to decrease the distance between segments, covering more ground and retrieving more detail in the representation of the space.

Some parts of the methods used in this thesis required defining threshold values. This is the case for the segmentation. Larger segments may collect necessary data, similarly an abundance of segments. If the segments are too short not all river space is captured, and if the segments are spaced out too much, some areas are not represented in the typology. Its a trade off between redundancy and insufficiency. The best possible values for the segmentation were not researched in this thesis, and here opportunities for future research lay.

This thesis uses a clustering algorithm to develop the typology, namely a k-means clustering algorithm. However, problems can arise when using such algorithms. Boongoen and Iam-On [2018] identified two major problems when applying clustering algorithms: selecting the appropriate technique and specifying the settings of the algorithm, such as the number of clusters in K-means clustering. Different techniques can cause different clusters, and changing parameter settings within one technique can reveal various structures in the data. As this thesis did not research the best clustering approach to utilize, the choice for clustering was not of major significance. However, as different algorithms and parameter settings may result in different resulting typologies, this is something to be taken into account when extending and reproducing the method. This thesis also pre-processed some of the metrics due to their large amount of outliers and skewness, and used a Yeo-Johnson power transformer to transform the input. These choices affect how the algorithm clusters the data. The value of k was chosen following the results of the clustergram and goodness-of-fit scores, but the choice of k is still done by the author. Future research could include the exploration of other clustering algorithms for river space typology development or changing the way k-means was implemented to see differences.

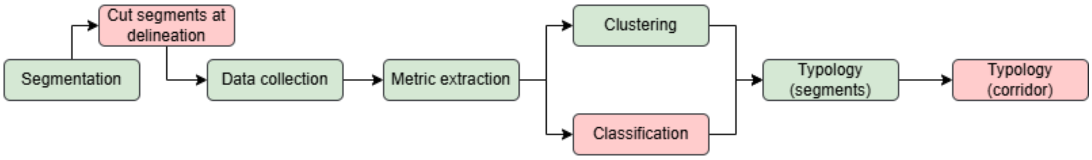


Figure 6.2.: Proposed workflow typology cutting segments at delineations, combining clustering and classification, and final corridor typology. Additional steps are shown in red.

An alternative to clustering could be a data-driven classification of the spaces. The clustering step in the typology development would be replaced or enhanced by defining classes of river spaces based on the metric distributions. For example, a space can have low, medium or high vegetation, combined with low, medium or high building cover. This was also done in the typology developed in this thesis, as main types were defined by these metrics. This was done as it was very difficult to visually differentiate between these subtypes. The shape of the terrain could be categorized as flat, sloped towards the river or curved. The clustering algorithm has a similar approach, except there is no control over the types made besides defining input, which gains in objectivity and may result in underlying patterns which are not as obvious. However, the typology may benefit from some extra input, guiding the typology or classification, or could benefit from supervised clustering instead of unsupervised.
Specifically in the case of this thesis, there were some metrics that had the median value at 0, influencing the cluster by being more of a binary indicator (is this metric present or not?). This heightens the influence of this metric. In these cases, a classification instead of clustering may be more applicable for these type of metrics.

Although skewness was reduced, some features still exhibited large peaks, particularly around specific values. This can distort distance calculations in K-Means, leading to poor cluster separation. An alternative approach could be K-Prototypes, which allows treating such features as categorical instead. By defining meaningful categories based on observed peaks (e.g., zero vs. non-zero or low/medium/high bins), we could improve cluster formation and interpretability.

When defining the types of river spaces following the clustering algorithm, it was noted that there is still quite some variety within a cluster. Ideally, this would be reduced to have more clear and meaningful descriptions. Improving the clustering or adding a classification, adding categorical metrics such as order of landuse or more specific numerical metrics, may lessen variety within types.

Lastly, the developed typology was compared to an existing neighbourhood typology. Segments were intersected with the polygon objects from this dataset, and it was investigated how clusters are distributed over these neighbourhoods types. This is a very limited way to validate the developed typology, but differences between clusters were visible, and similarities between subtypes could be seen. Not much can be said about the quality of the developed typology by comparing it to this existing typology, as they describe types of different things, river spaces and neighbourhoods. Nevertheless, it was interesting to see the correlation and the differences between the two.

7. Conclusion

The objectives of this thesis were the delineation of the urban river space, the semantic enrichment of the description of the space, and the development of a data-driven typology for it.

The first research question asked was 'How can urban river spaces be delineated and analyzed using (2D) and 3D spatial data?'. Three delineation methods were proposed; one for the first line of buildings, one for visible space, and one for floodable space. These methods were chosen following the literature review and its relevance to use cases. The existing 2D delineation method used the first line of buildings, and this approach was adapted by providing a method to compute this line from building data automatically. Visible space was delineated as a viewshed from the riverline. Floodable space was delineated using the 100 year flood data. These were the solutions offered in response to the existing 2D delineation method, and the question of what spaces can be distinguished as river space. Shortcomings of the proposed methods were described in the results section, and their potential for future research in the discussion.

The second research question asked was 'What types of spatial configurations and qualities can be observed in urban river spaces as seen through 3D delineation and spatial analysis?'. To be able to answer this, it was important to understand how river spaces differ and are characterized. The literature review provided an overview of the characteristics of the urban river space, together with the described use-cases, which gives more of an applicable description. Fundamental indicators were defined as the spatial characteristics of the river space, flooding potential considerations, biological components and the built environment, and the visibility from the river. A method for data-driven river space typology was proposed using cross-sectional segmentation and k-means clustering. The cross-section segmentation followed from the literature review, as research often described the space by its cross-section. Data regarding the fundamental indicators was collected along these cross-sections, and metric values were computed per segments. The metric values were used as input in the k-means clustering algorithm. 10 clusters were derived from almost 150 000 segments and interpreted to obtain a river space typology of the Netherlands. The types were visualized to serve as a tool for descriptive applications, or to serve as a decision making tool to identify the interventions needed to transform a space into the ideal state.

A. Reproducibility self-assessment

A.1. Marks for each of the criteria

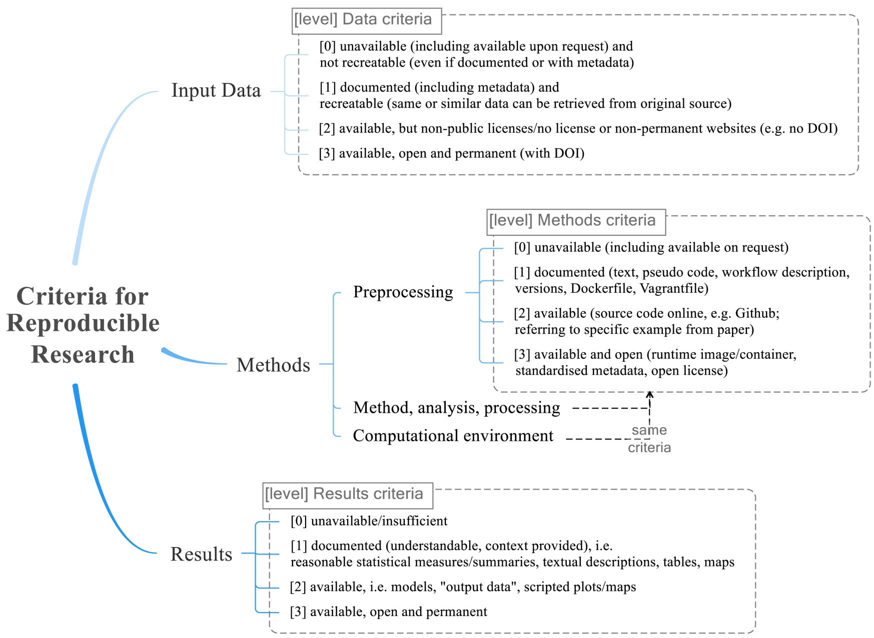


Figure A.1.: Reproducibility criteria to be assessed.

Grade/evaluate yourself for the 5 criteria (giving 0/1/2/3 for each):

1. input data: 3.
2. preprocessing: 2
3. methods: 2
4. computational environment: 3
5. results: 2

A.2. Self-reflection

To enhance reproducibility of the results of this master thesis, there was a requirement to only use open data and open source software. Datasets used in this thesis were all openly

available via various organizations, as described in table 4.1. Therefore, the same data can be used for other research or to validate results. The steps taken to prepare the data for processing are described in this thesis (chapter 4), detailing each step and providing pseudocode for the main algorithms. Data specifications were described, and quality was discussed. Most datasets are permanently available with DOI, but for some derived datasets this was more of a problem. The Github repository contains details on the specific functions used, as data was prepared using QGIS and python code. Every step of the process done in python is given and explained in the repository. The code used to compute segments, parameter values and metric values was documented and is available on GitHub. The requirements file specifies packages to be installed.
Pre-processing of data was done in the open source software QGIS using python code. Some data retrieval was done manually, and some automatically. The scripts to do this are provided in the GitHub, but they need to be executed separately, not everything can be done in one go automatically. The segmentation script can be executed in one go, steps are described in the code, pseudocode is provided in this thesis, and steps are detailed. Parameter extraction is described and script is given, same for metric extraction and the clustering. The delineation script require some specific input, described and defined in the code. When randomly sampling the segments, a seed was used in order to enhance reproducibility. This can be found back in the GitHub repository. Similarly, the k-means clustering algorithm uses a seed to ensure the same results when running the script.

B. Landuse: Basisregistratie Grootschalige Topografie dataset

B.1. Relevant layers

The BGT (*Basisregistratie Grootschalige Topografie*, Large-Scale Topography Basic Registry) is the detailed digital basis map of the Netherlands, where the location of a range of physical objects is established. The BGT is divided into 35 layers (`FeatureTypes`) and one extra layer called *'plaatsbepalingspunt'* (positioning point), describing the boundaries between objects. The `FeatureTypes` used in this thesis are shown in table B.1. These are the layers of the dataset relevant to fundamental characteristics of urban river spaces, including land cover such as vegetation and water, and functional elements such as roads and buildings. Some layers in the dataset have mandatory sub-categories defined in one of the columns which identifies the type of the object. Table B.1 shows these sub-categories and which column identifies them. These are used to distinguish between things like forest and grass, or pedestrian path and road.

B.2. Complete dataset

The complete list of `FeatureTypes` is as follows: *'bak'*;bin, *'begroeidterreindeel'*;vegetated terrain, *'buurt'*;neighbourhood (topographic), *'functioneelgebied'*;functional area, *'gebouwinstallatie'*;building installation, *'installatie'*;installation, *'kast'*;box, *'kunstwerkdeel'*;infrastructure piece, *'mast'*;mast, *'onbegroeidterreindeel'*;non-vegetated terrain, *'ondersteunendwaterdeel'*;areas serving a supporting role to water, *'ondersteunendwegdeel'*;supporting road parts, *'ongeclassificeerdobject'*;unclassified object, *'openbareruimte'*;public space, *'openbareruimtelabel'*;public space label, *'overbruggingsdeel'*;bridge part, *'overigbouwerk'*; other structure, *'overigescheiding'*; other barrier, *'paal'*;pole, *'pand'*;building, *'put'*;well, *'scheiding'*;barrier, *'sensor'*;sensor, *'spoor'*;track, *'stadsdeel'*;city part, *'straatmeubilair'*;urban furniture, *'tunneldeel'*;tunnel part, *'vegetatieobject'*; vegetation object, *'waterdeel'*;waterpart, *'waterinrichtingselement'*;water feature element, *'waterschap'*;water management area, *'wegdeel'*;roadpart, *'weginrichtingselement'*;road feature element, *'wijk'*;district (social geographic).

FeatureType	**Description**	**Sub-categories**
waterdeel	Water plain polygons	`bgt-type`: waterloop (waterways), watervlakte (waterplain), zee (sea), and greppel, droge sloot (ditches)
ondersteunendwaterdeel	Polygon objects of areas serving a supporting role to the water plains.	`bgt-type`: oever, slootkant (embankment) and slik (silt). Embankment is an area in direct contact with water, including the area between highwater and lowwater line. Silt is a non-vegetated ground which floods almost every highwater
begroeidterreindeel	Polygon objects of terrain covered in vegetation.	`FysiekVoorkomen`: loofbos, gemengd bos, naaldbos, heide, struiken, houtwal, duin, grassland overig, moeras, rietland, kwelder, fruitteelt, boomteelt, bouwland, grassland agrarisch, groenvoorziening
vegetatieobject	Points of vegetation objects, such as trees or hedges.	`punt: plus_type` : boom
onbegroeidterreindeel	Polygon objects of terrain covered in non-vegetation	`FysiekVoorkomen`: erf, gesloten verharding, open verharding, half verhard, onverhard, zand
wegdeel	Polygon objects of roads and paths	`Functie`: OV-baan, overweg, spoorbaan, baan voor vliegverkeer, rijbaan: autosnelweg, rijbaan: autoweg, rijbaan: regionale weg, rijbaan: lokale weg, fietspad, voetpad, voetpad op trap, ruiterpad, parkeervlak, voetgangersgebied, inrit, woonerf
ondersteunendwegdeel	Polygon objects of areas serving a supporting role to the roads	*Explain*
pand	Polygon layer of building footprints	-

Table B.1.: Used `FeatureTypes` from BGT dataset with subcategories.

C. Metrics

C.1. Distributions

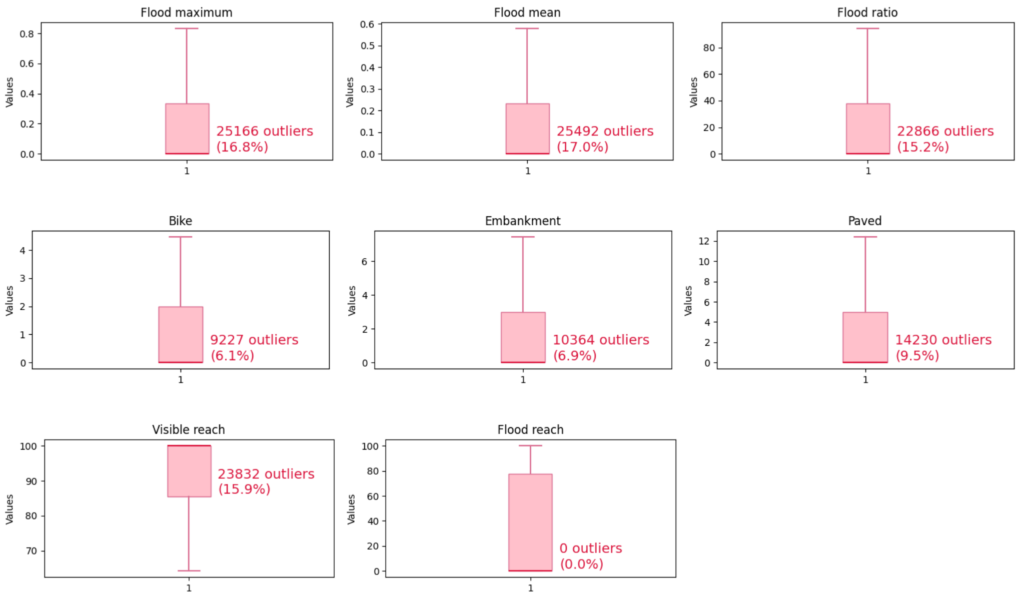


Figure C.1.: Boxplots for metrics with median at 0 or 100.

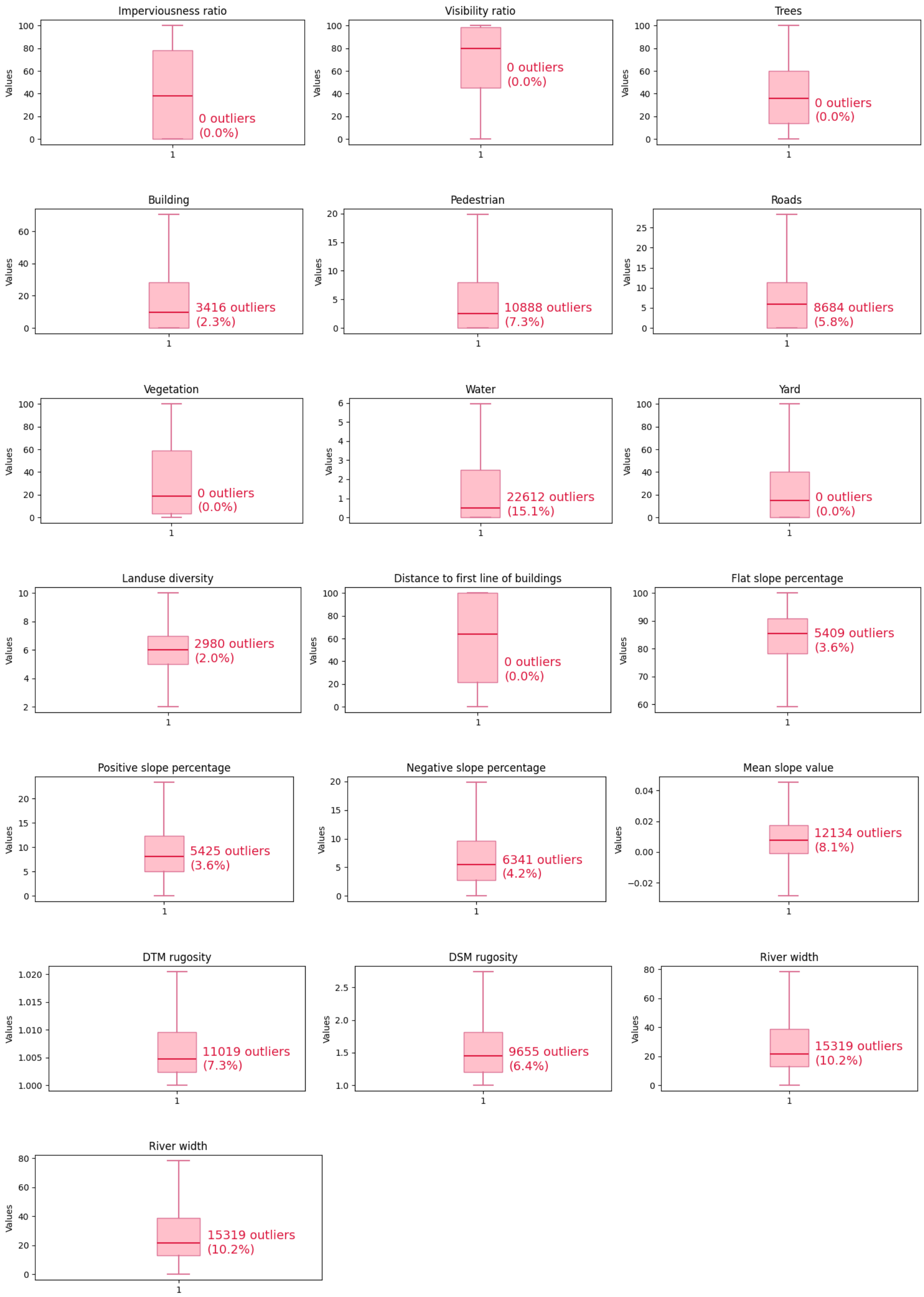


Figure C.2.: Boxplots for metrics

C.2. Skewness

The table C.1 shows the skewness value of each metric befor and after applying Yeo-Johnson transform. Skewness is a measure of asymmetry of a distribution. Negative skew implies the tail is longer towards the left side, positive skew implies the tail is longer towards the right side. Reducing skewness can improve k-means clustering. If skewness is between -0.5 and 0.5, the data is fairly symmetrical, for skewness between $[-1, -0.5]$ and $[0.5, 1]$, the data is moderately skewed. High skewness is considered lower than -1 or higher than 1. Yeo-Johnson transformation puts almost all metric distributions in the fairly symmetrical category, except for visible space ratio, water and mean slope, which have values slightly below -0.5 or above 0.5. Two metrics actually get more skewed after the Yeo-Johnson transform. Therefore, these will not be transformed. Their skewness values are also in bounds before transforming.

Metric	Before	After
Flood ratio	1.26	0.73
Flood max	4.45	1.04
Flood mean	4.00	1.07
Bike	8.53	0.97
Embankment	7.21	0.37
Paved	4.35	0.37
Imperviousness ratio	0.2372	-0.3373
Visible space ratio	-0.6899	-0.5472
Trees	0.3605	-0.3276
Building	1.2389	-0.0259
Pedestrian	3.4344	0.0994
Roads	2.9466	-0.0317
Vegetation	0.7521	-0.1480
Water	3.5156	0.5406
Yard	0.9182	-0.1144
Landuse diversity	-0.1203	-0.0277
First building distance	-0.1687	-0.3354
Flat surfaces	-2.5814	-0.4131
Positive slope	1.8173	0.0189
Negative slope	1.5495	-0.0190
Mean slope	2.8766	-0.5072
DTM rugosity	9.1684	0.0000
DSM rugosity	2.9074	0.1479
River width	4.167	-0.0142

Table C.1.: Skewness values before and after Yeo-Johnson transform

Even though skewness is reduced, many of the histograms still show large peaks due to many segments having the same (often 0) value for that metric. The transform maps metrics for which the distribution has more continuity better to a more Gaussian distribution (such as rugosity metrics and riverwidth) than the metrics with 1 big bin (such as roads or building). Nevertheless, the transform reduces the impact of these as the distance between the peak and the rest of the distribution is reduced. The histograms in figure C.4 show the metric distributions after transformation (except the blue histograms, which were not transformed).


Histogram with KDE of flood_rat
flood_rat
Histogram with KDE of bike
bike
Histogram with KDE of embankment
embankment
Histogram with KDE of paved
paved
Histogram with KDE of imperv_rat
imperv_rat
Histogram with KDE of vis_rat
vis_rat
Histogram with KDE of trees
trees
Histogram with KDE of building
building
Histogram with KDE of pedestrian
pedestrian
Histogram with KDE of roads
roads
Histogram with KDE of vegetation
vegetation
Histogram with KDE of water
water
Histogram with KDE of yard
yard
Histogram with KDE of land_div
land_div
Histogram with KDE of building_line
building_line
Count

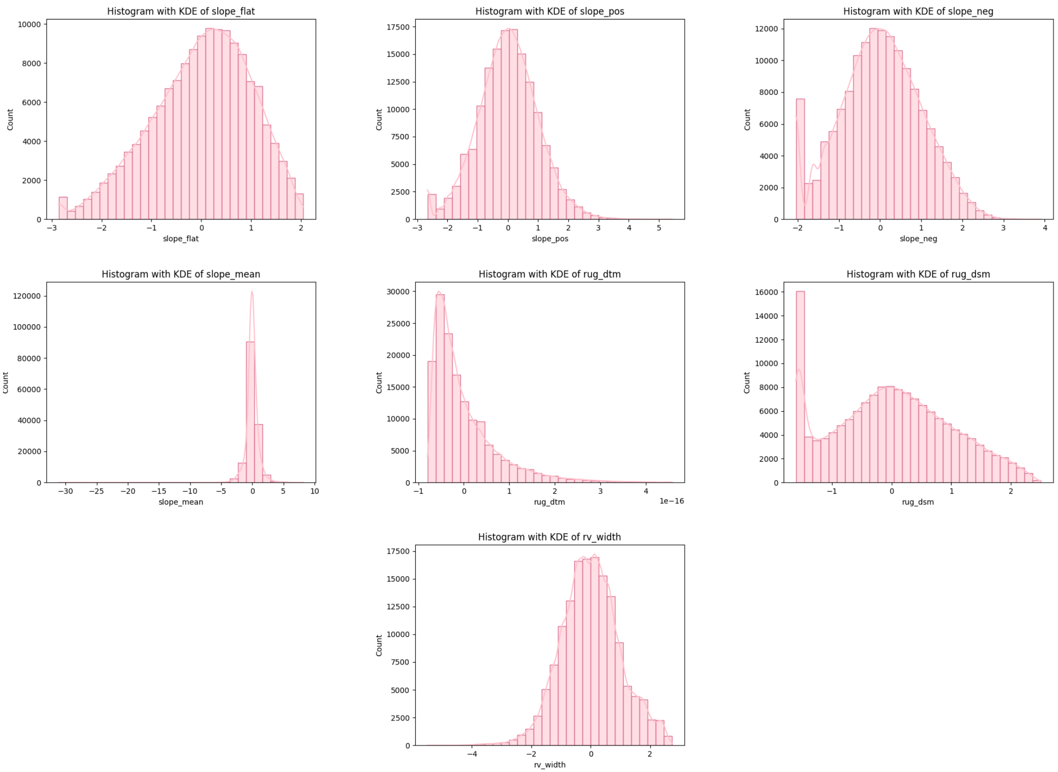


Figure C.4.: Histograms for metrics after Yeo-Johnson transform

D. Clusters

This appendix contains the mean and median values per metric per cluster. The boxplots of each metric per cluster are provided, visualizing differences between clusters and the distributions within each cluster.

Table D.1.: Median and mean values per cluster (Part 1)

Cluster	Flood ratio		Imperviousness ratio		Visibility ratio		Trees	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
0	0	14.51	71.29	71.16	73.76	67.72	37.13	37.08
1	0	35.28	0	1.82	89.11	72.34	34.65	40.39
2	0	17.01	87.13	85.93	83.17	68.86	22.28	27.43
3	0	21.61	9.90	9.47	84.65	72.76	56.93	52.53
4	0	16.30	41.58	41.69	78.71	70.22	49.01	47.15
5	0	22.12	17.82	15.94	77.23	68.81	58.42	54.85
6	0	14.89	91.09	90.30	72.77	67.45	21.78	24.65
7	0	21.16	38.12	39.10	83.66	70.48	46.04	44.73
8	0	12.66	100.00	95.75	71.78	66.96	20.30	22.77
9	0	19.06	53.96	53.75	75.25	67.71	43.56	42.87

Table D.2.: Median and mean values per cluster (Part 2)

Cluster	Building		Pedestrian		Bike		Embankment	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
0	31.68	33.94	5.94	8.21	0	0.81	0	1.11
1	0	0.38	0	2.16	0	1.35	1.98	3.84
2	0	3.25	4.46	13.42	0	2.08	0	1.12
3	15.35	18.75	1.98	4.10	0	0.76	0.99	2.06
4	22.77	25.04	3.96	6.26	0	0.85	0	1.38
5	12.87	14.71	2.48	4.66	0	1.34	1.49	2.50
6	29.21	31.02	6.44	10.78	0	1.35	0	0.96
7	0	2.52	2.48	7.11	0	2.38	0.50	2.54
8	43.07	44.86	7.43	9.94	0	0.67	0	0.39
9	20.30	22.31	4.95	7.07	0	1.38	0.50	1.94

Table D.3.: Median and mean values per cluster (Part 3)

Cluster	Paved		Roads		Vegetation		Water	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
0	2.48	4.89	7.43	9.71	5.94	8.69	0.50	1.20
1	0	1.86	0	4.05	79.70	69.74	1.49	5.68
2	3.47	11.92	10.89	20.71	7.43	13.40	0	1.39
3	0	2.56	5.45	6.80	14.36	21.13	0.50	3.97
4	1.49	3.81	6.44	8.46	9.90	14.19	0.50	2.25
5	0	2.56	5.45	7.10	28.71	29.19	0.50	3.43
6	3.47	7.21	9.41	11.91	4.95	7.68	0	0.88
7	1.98	7.17	10.40	15.77	34.16	34.55	0.50	3.85
8	2.97	5.65	7.92	10.36	0.50	3.30	0	0.69
9	1.98	4.69	6.93	9.55	19.80	20.20	0.50	2.35

Table D.4.: Median and mean values per cluster (Part 4)

Cluster	Yard		Landuse diversity		Distance to first line of buildings		Flat slope percentage	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
0	30.20	31.04	7.00	6.65	18.91	18.85	86.23	85.14
1	0	9.14	5.00	4.90	100.00	99.47	83.14	80.98
2	9.41	30.02	6.00	5.53	100.00	95.79	90.32	88.03
3	39.11	39.17	7.00	6.50	20.90	21.31	83.87	82.45
4	37.62	37.18	7.00	6.77	18.91	19.20	85.03	83.84
5	30.20	33.64	7.00	6.97	60.70	60.56	81.63	80.15
6	24.75	27.40	7.00	6.55	43.78	45.75	88.39	86.98
7	5.94	21.51	6.00	6.42	100.00	96.61	83.25	81.27
8	21.29	23.97	6.00	6.03	12.44	13.13	89.29	88.21
9	27.23	29.76	7.00	7.16	50.75	52.38	83.61	81.92

Table D.5.: Median and mean values per cluster (Part 5)

Cluster	Positive slope percentage		Negative slope percentage		Mean slope value		DTM rugosity	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
0	8.11	9.07	4.86	5.79	0.01	0.01	1.00	1.01
1	8.94	10.45	7.10	8.57	5.90e-03	6.80e-03	1.01	1.01
2	5.97	7.20	3.37	4.77	8.00e-03	0.01	1.00	1.01
3	8.75	9.73	6.67	7.81	6.60e-03	7.02e-03	1.01	1.01
4	8.39	9.21	5.84	6.95	8.20e-03	8.48e-03	1.00	1.01
5	10.19	11.14	7.14	8.71	9.60e-03	9.24e-03	1.01	1.01
6	7.05	8.16	3.95	4.85	0.01	0.01	1.00	1.01
7	9.14	10.55	6.63	8.18	8.50e-03	9.09e-03	1.01	1.01
8	6.25	7.21	3.88	4.58	8.20e-03	0.01	1.00	1.01
9	9.58	10.61	5.88	7.47	0.01	0.01	1.01	1.01

Table D.6.: Median and mean values per cluster (Part 6)

Cluster	DSM rugosity		River width	
	Median	Mean	Median	Mean
0	1.53	1.62	19.66	31.76
1	1.29	1.73	22.36	42.44
2	1.23	1.33	26.61	50.45
3	1.57	1.72	18.53	28.75
4	1.55	1.66	18.31	29.28
5	1.63	1.81	21.98	35.87
6	1.41	1.49	26.26	48.17
7	1.40	1.57	23.66	42.84
8	1.49	1.56	20.63	35.32
9	1.51	1.64	22.67	37.46

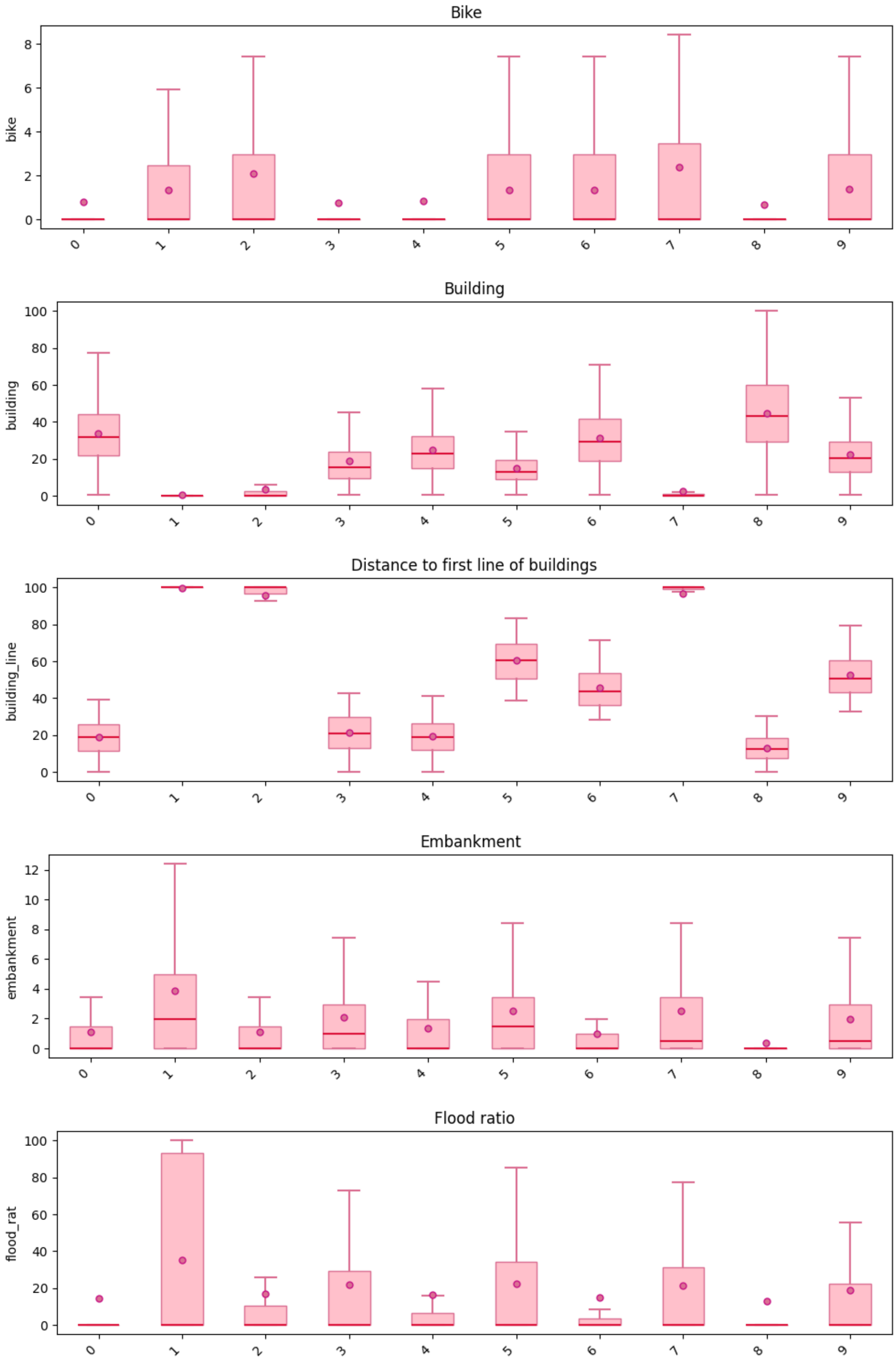


Figure D.1.: Boxplots for clusters metric values (Part 1)

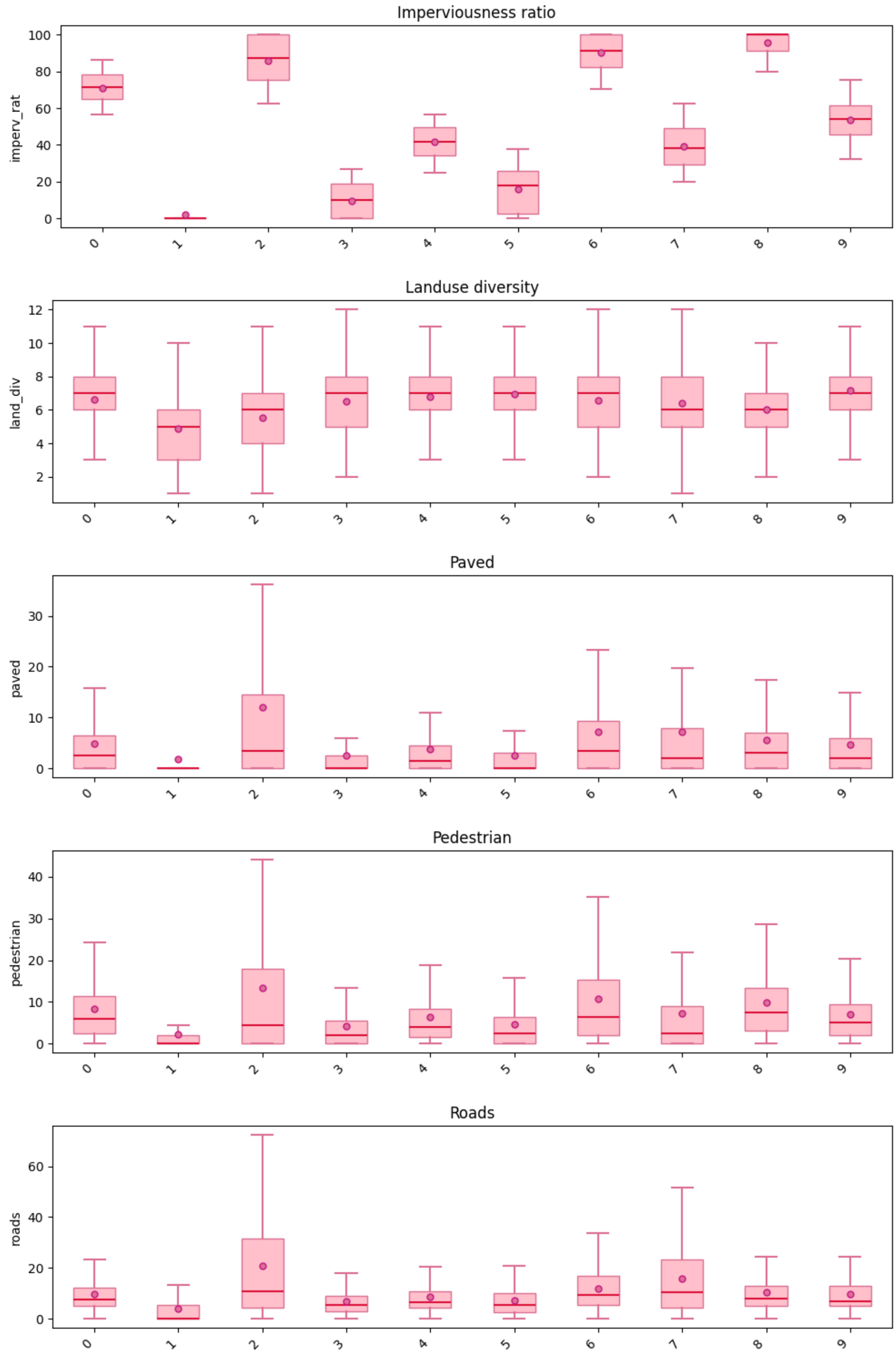


Figure D.2.: Boxplots for clusters metric values (Part 2)

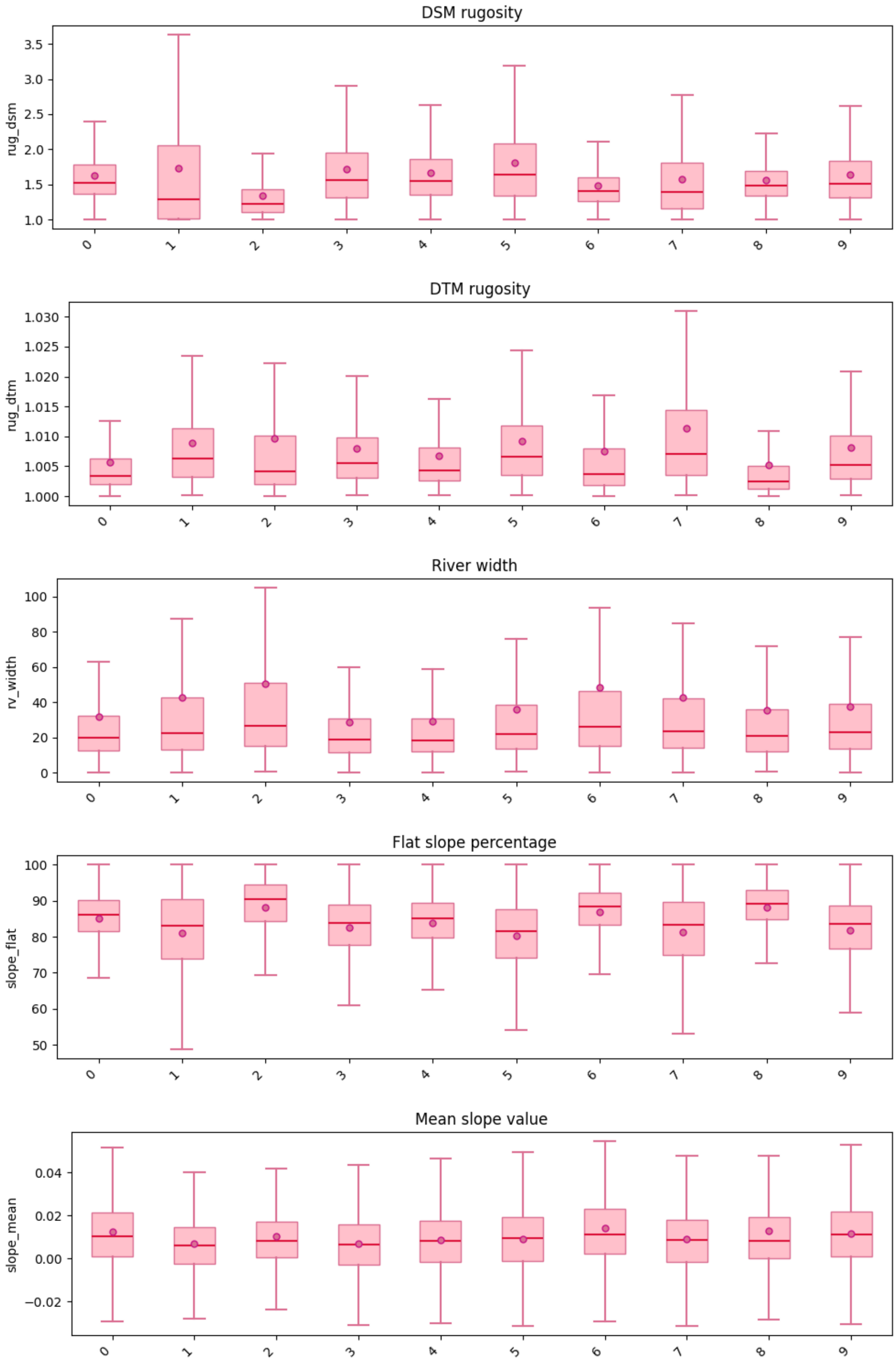


Figure D.3.: Boxplots for clusters metric values (Part 3)

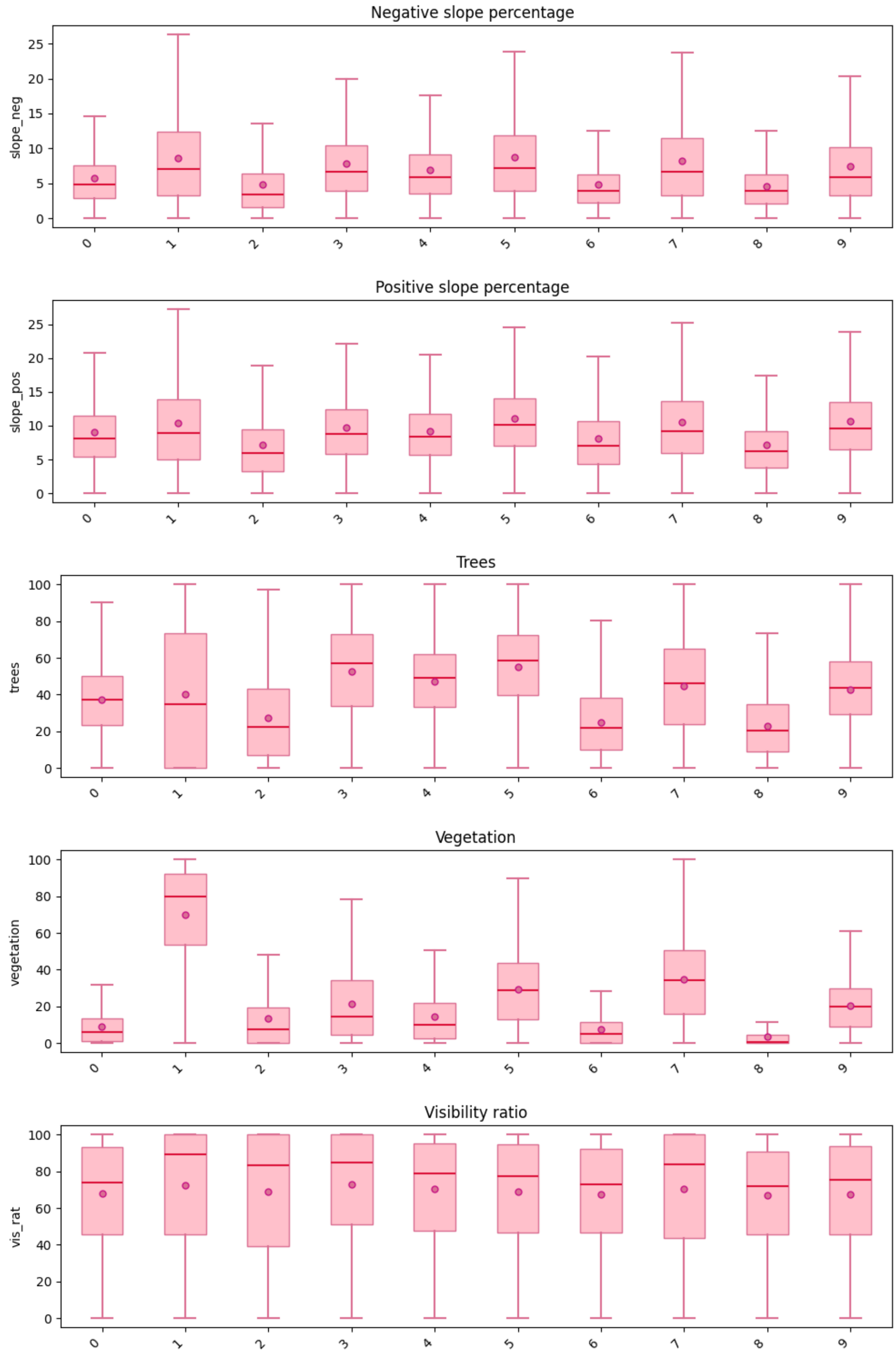


Figure D.4.: Boxplots for clusters metric values (Part 4)

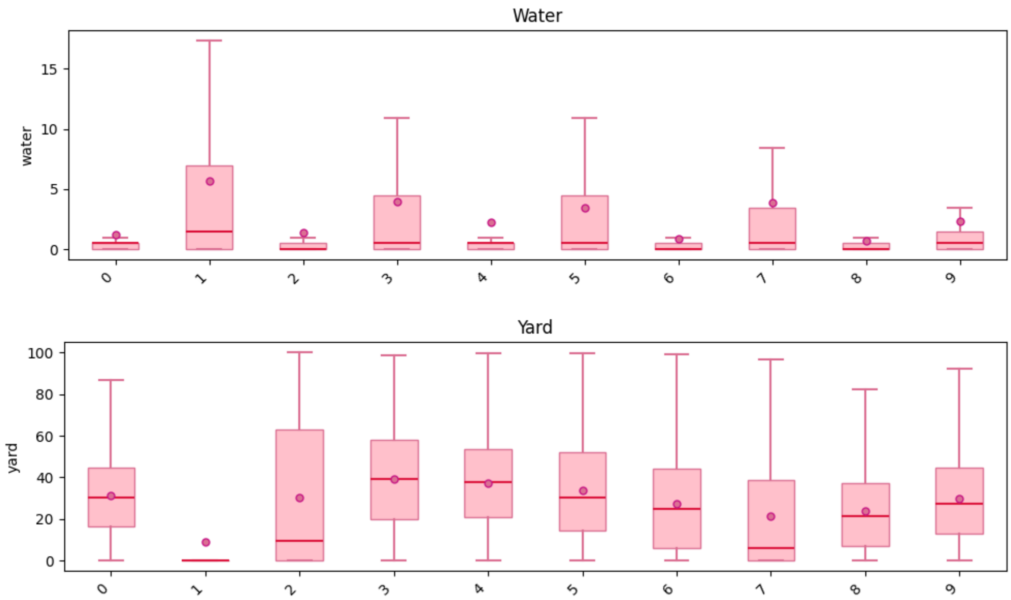


Figure D.5.: Boxplots for clusters metric values (Part 5)

E. Sampled segments

This appendix shows the parameter values and the orthophoto of 9 randomly selected segments for all 13 clusters. All the parameter plots follow the legend placed in figure E.1.

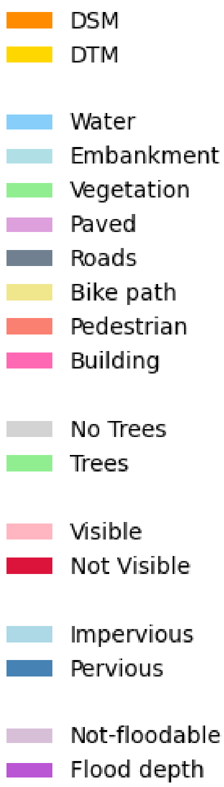


Figure E.1.: Legend

Figure E.2.: Sampled segments for cluster 0

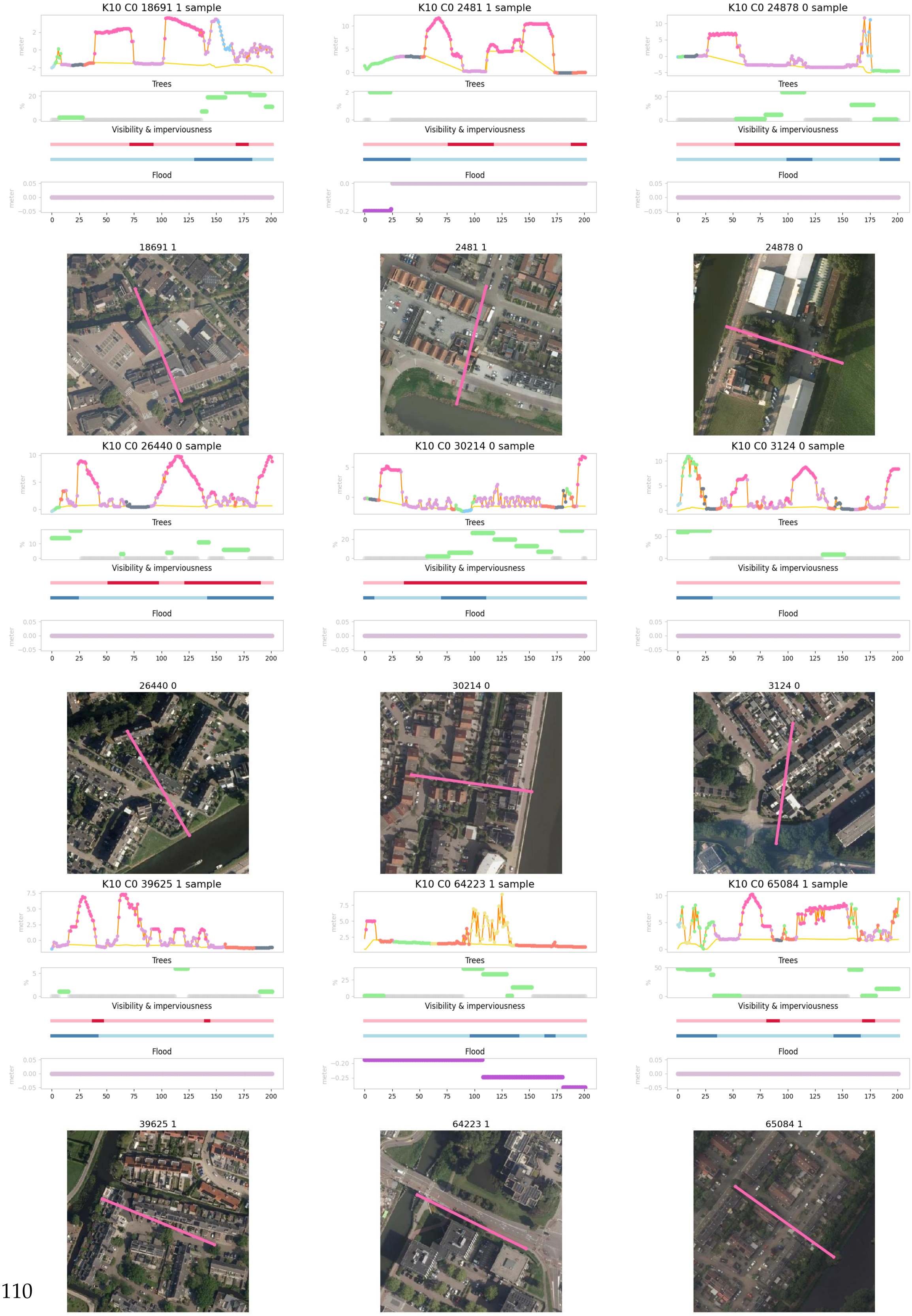


Figure E.3.: Sampled segments for cluster 1

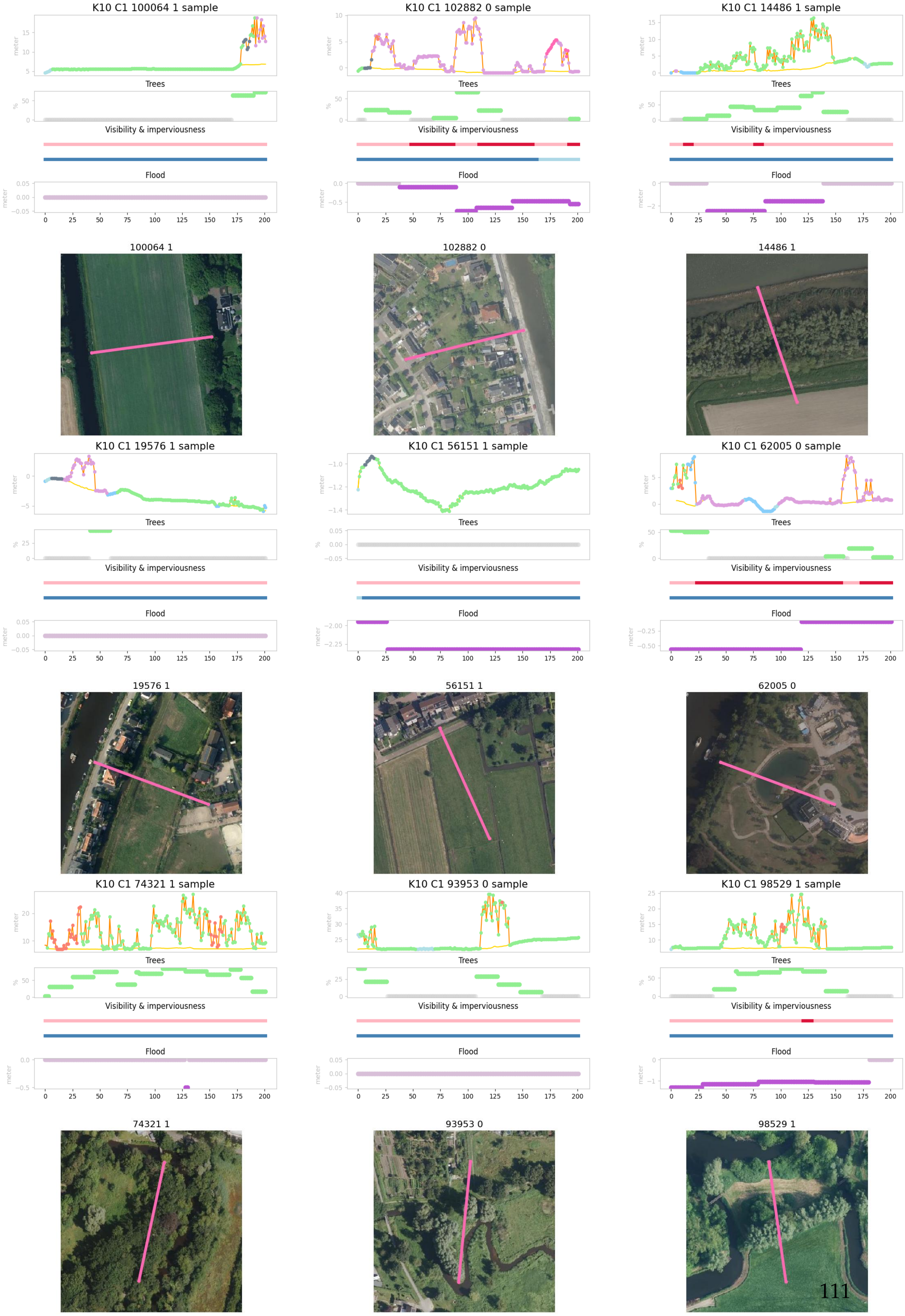


Figure E.4.: Sampled segments for cluster 2

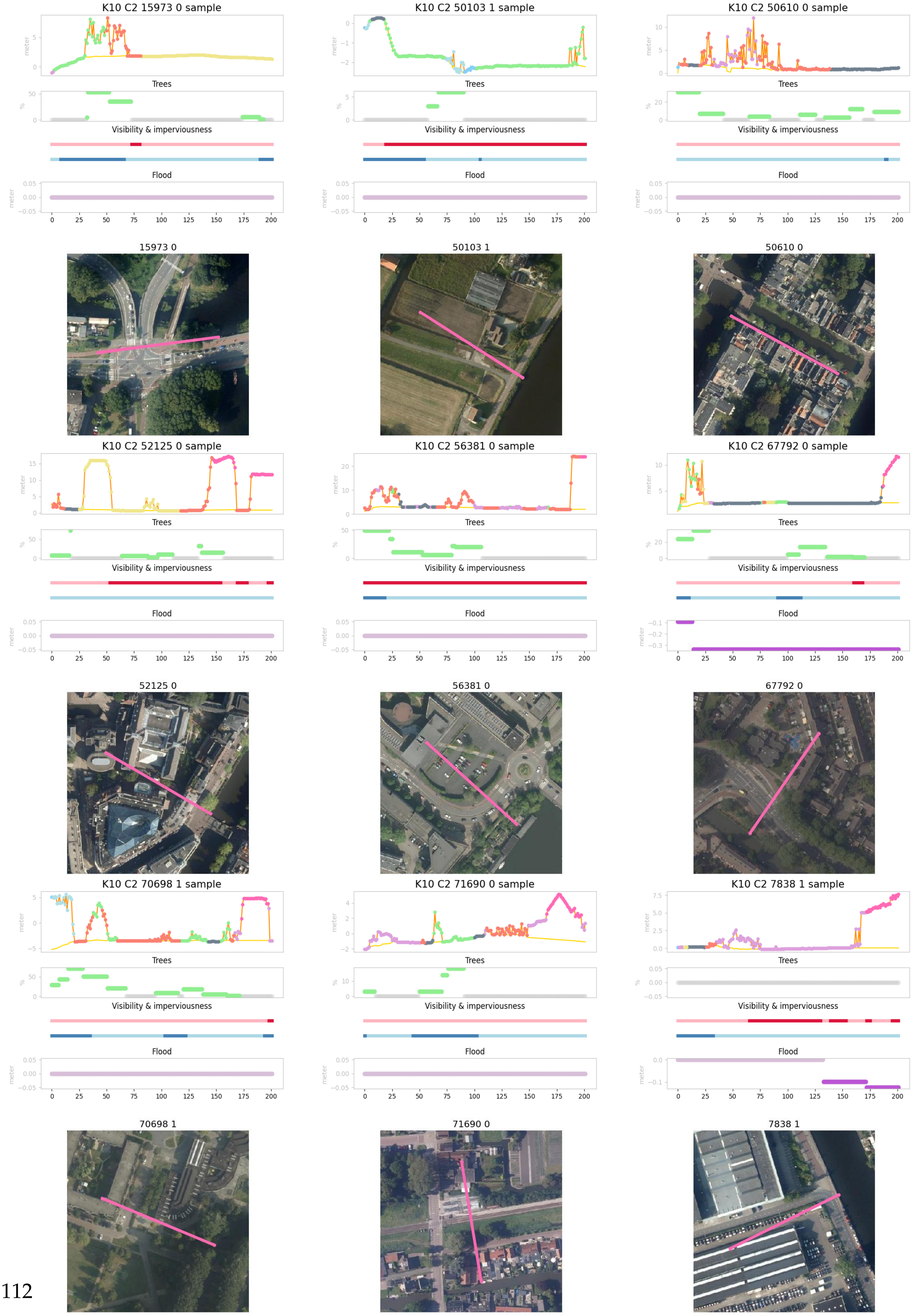


Figure E.5.: Sampled segments for cluster 3

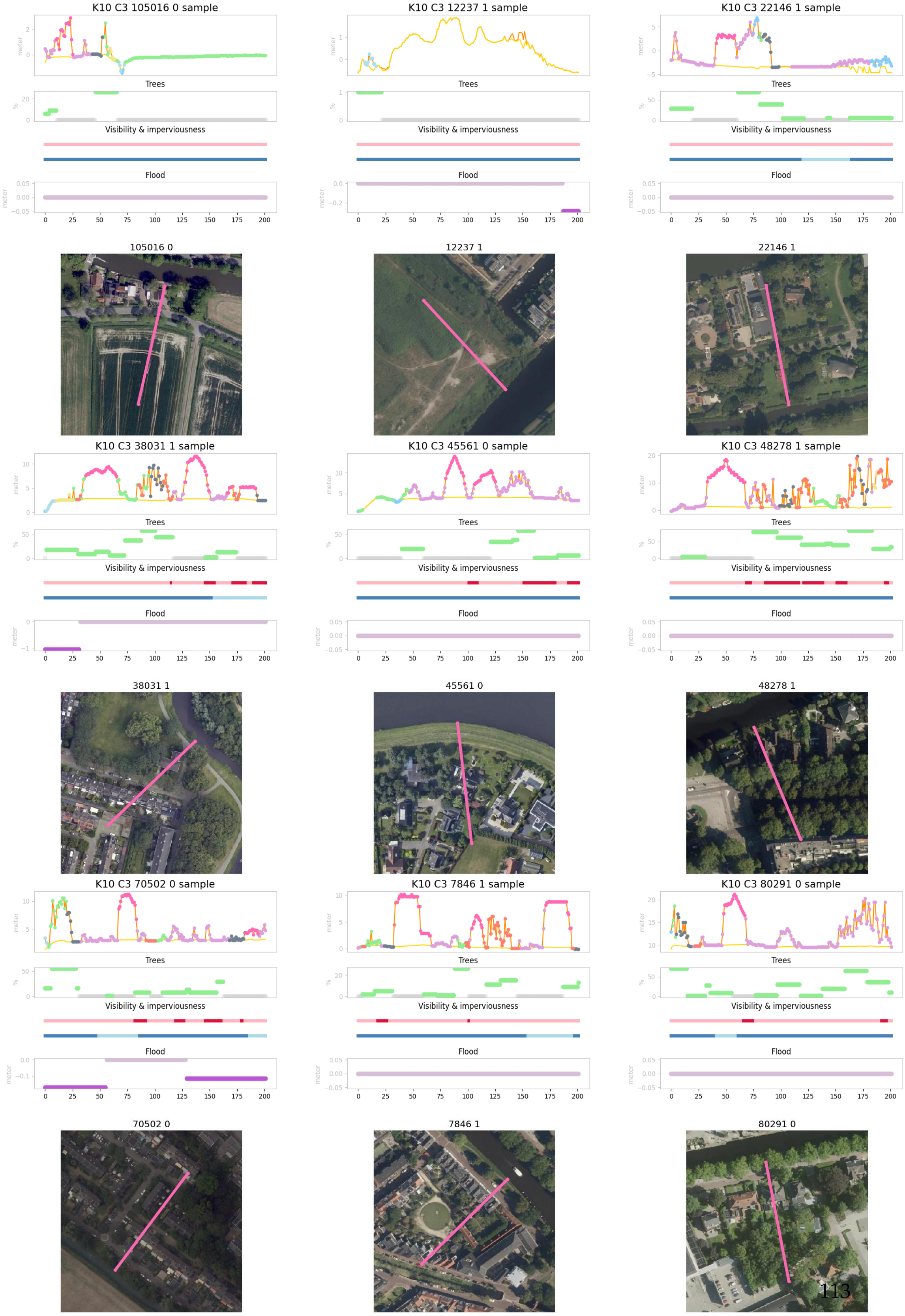


Figure E.6.: Sampled segments for cluster 4

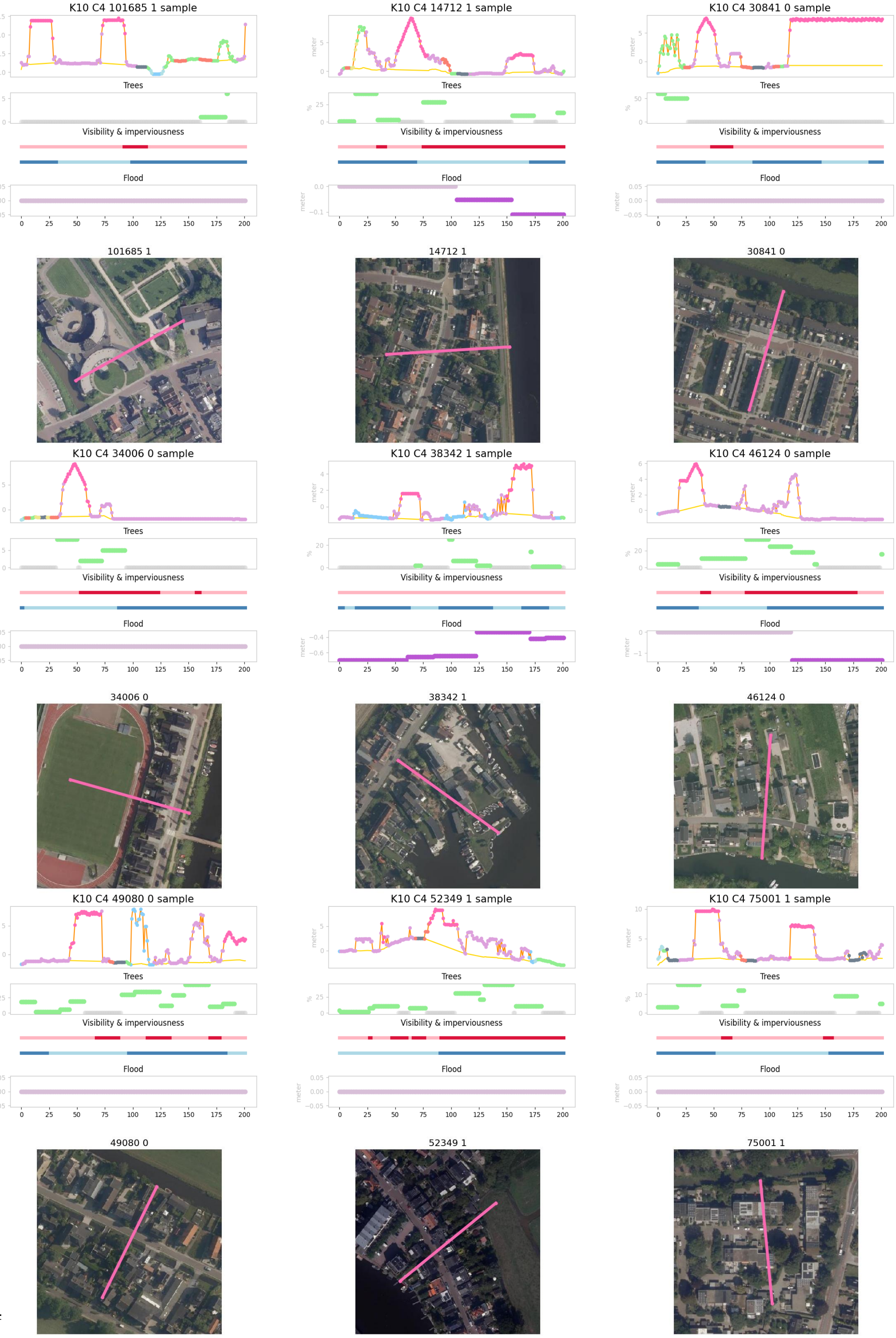


Figure E.7.: Sampled segments for cluster 5

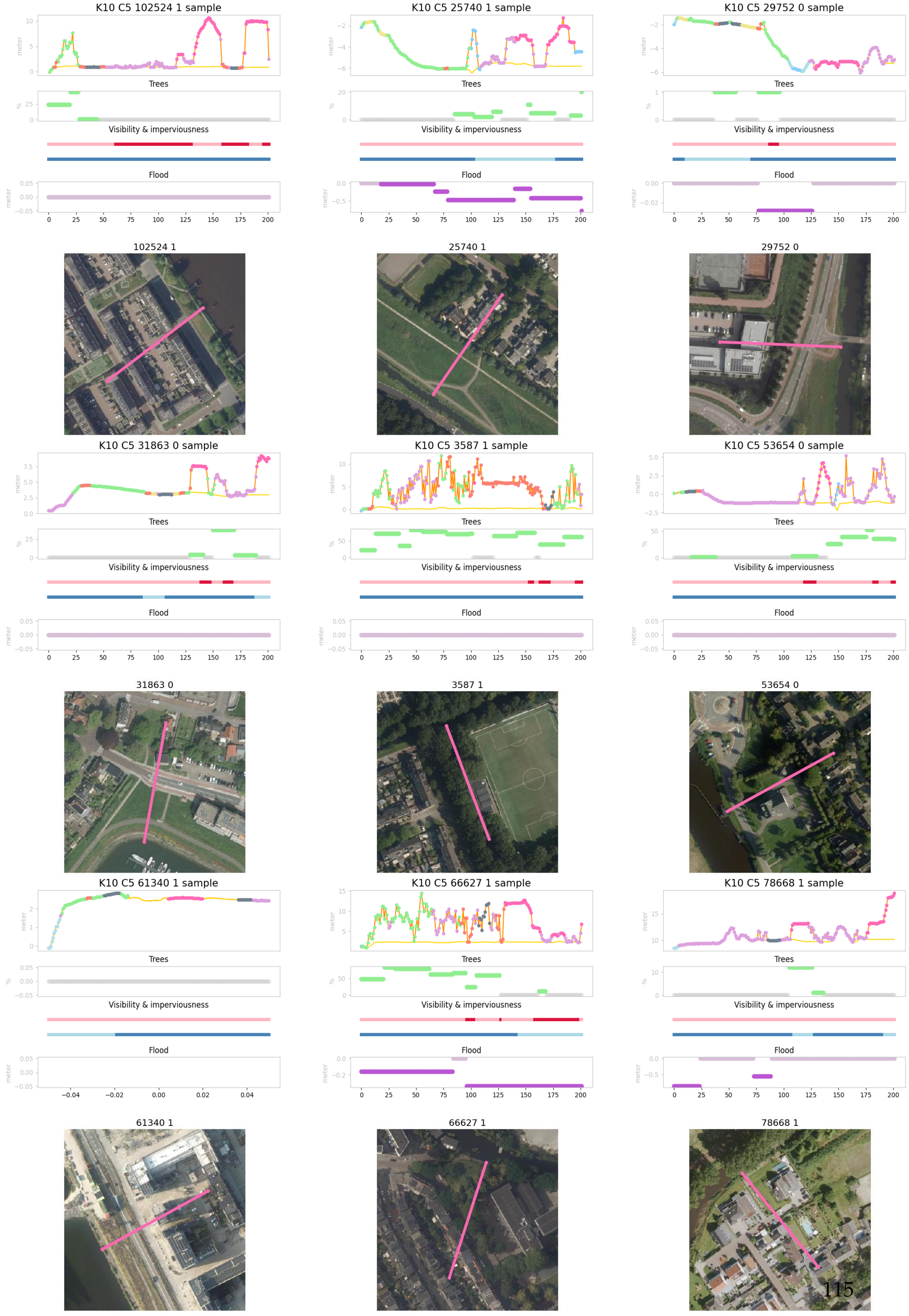


Figure E.8.: Sampled segments for cluster 6

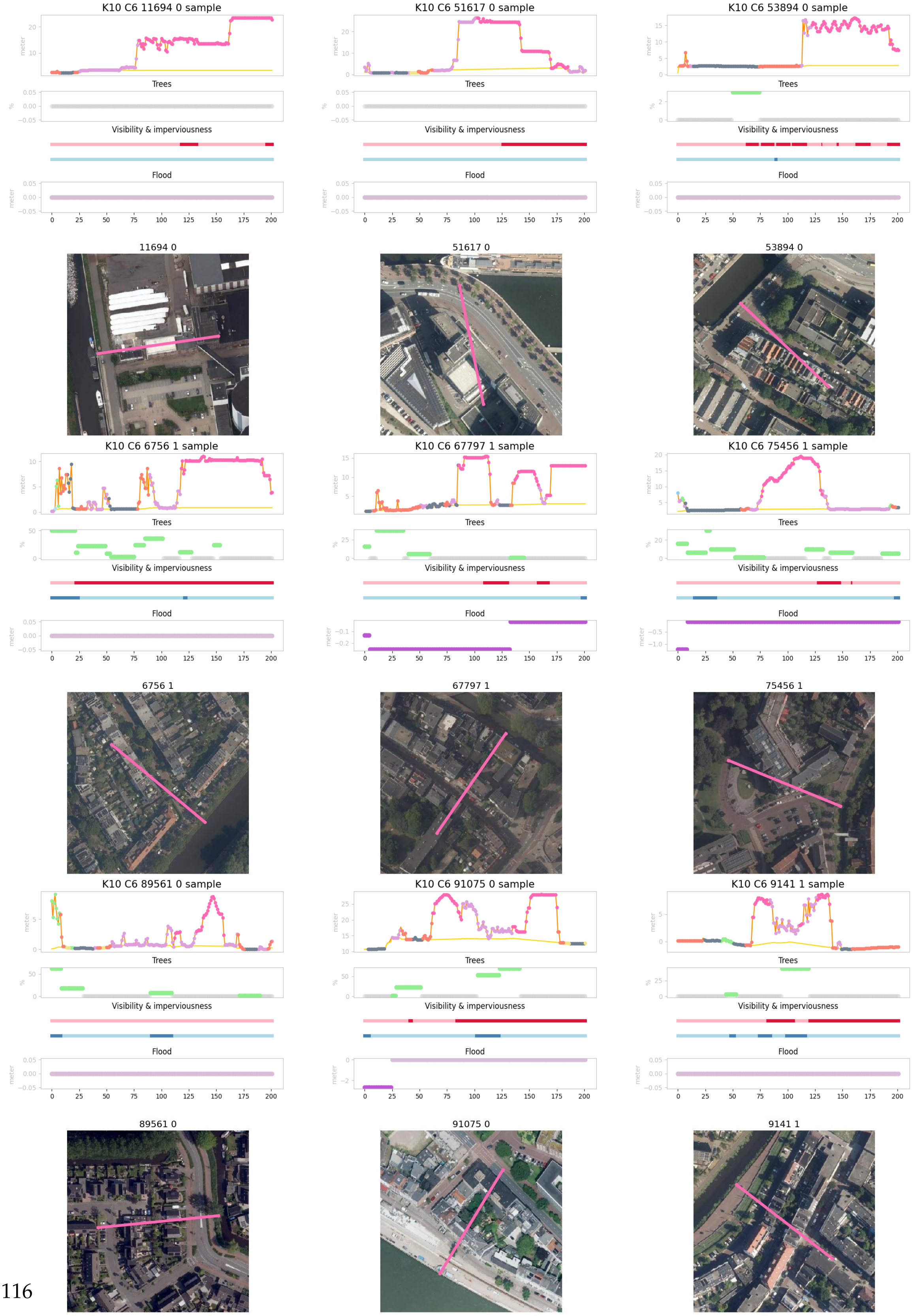


Figure E.9.: Sampled segments for cluster 7

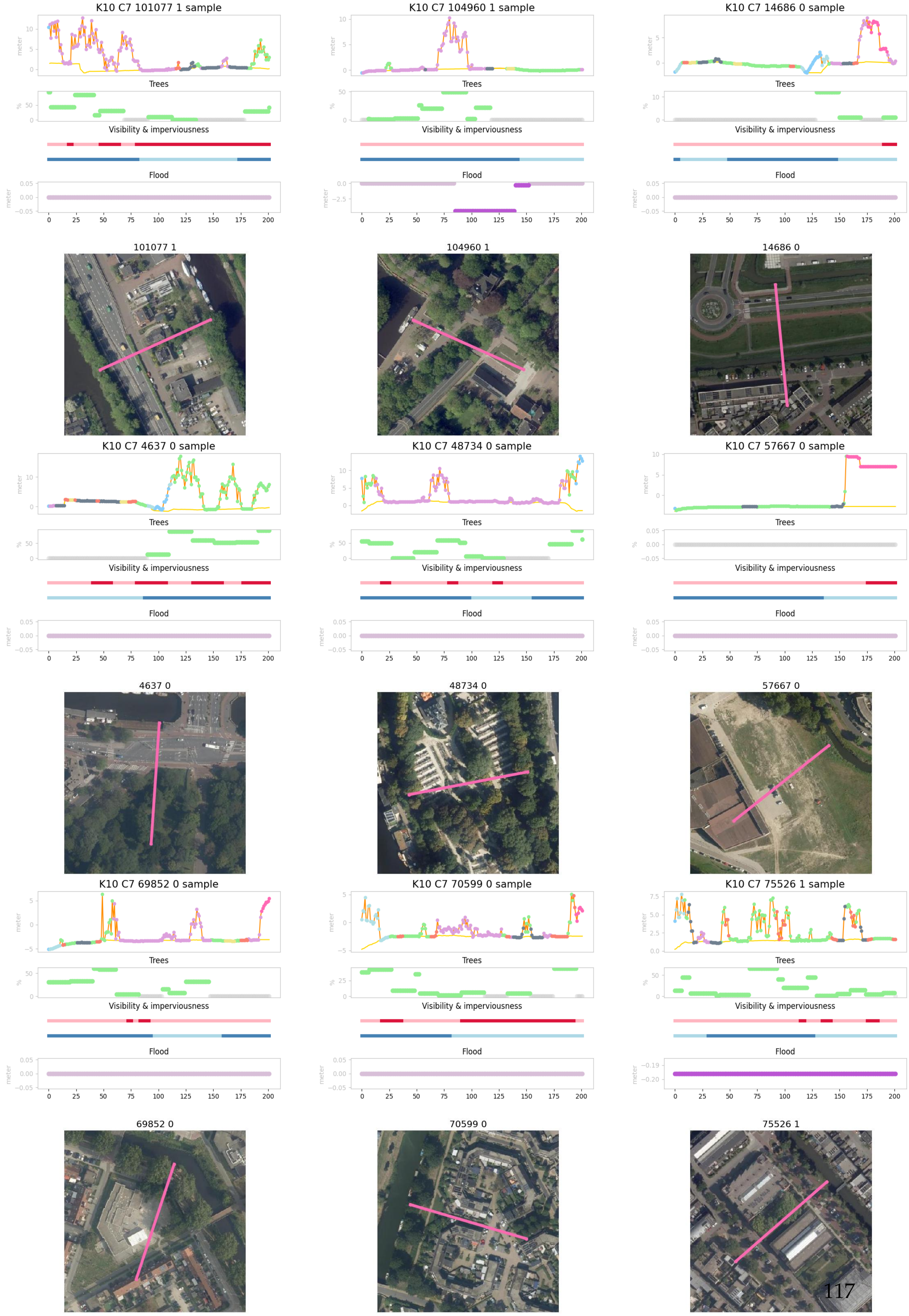


Figure E.10.: Sampled segments for cluster 8

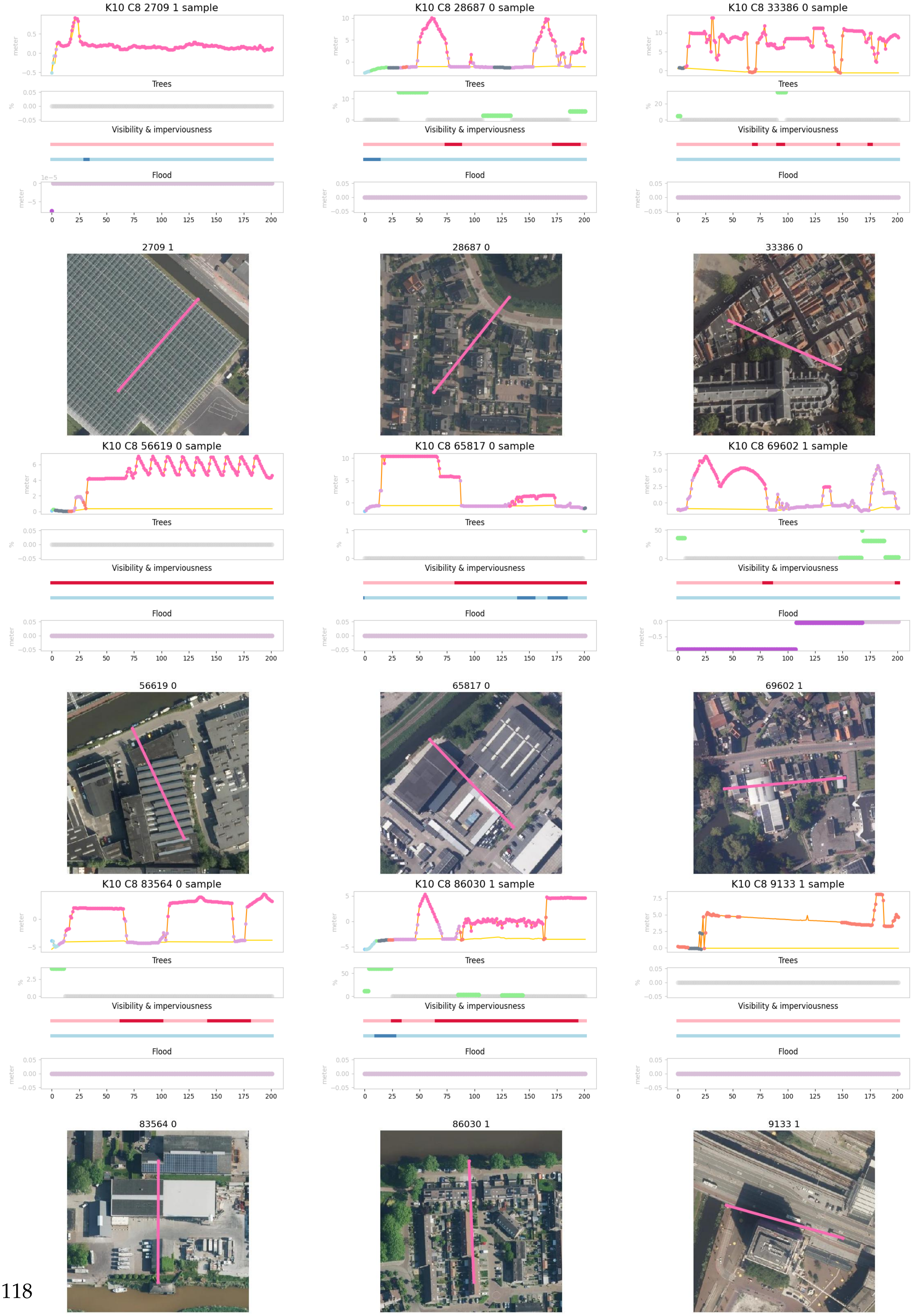
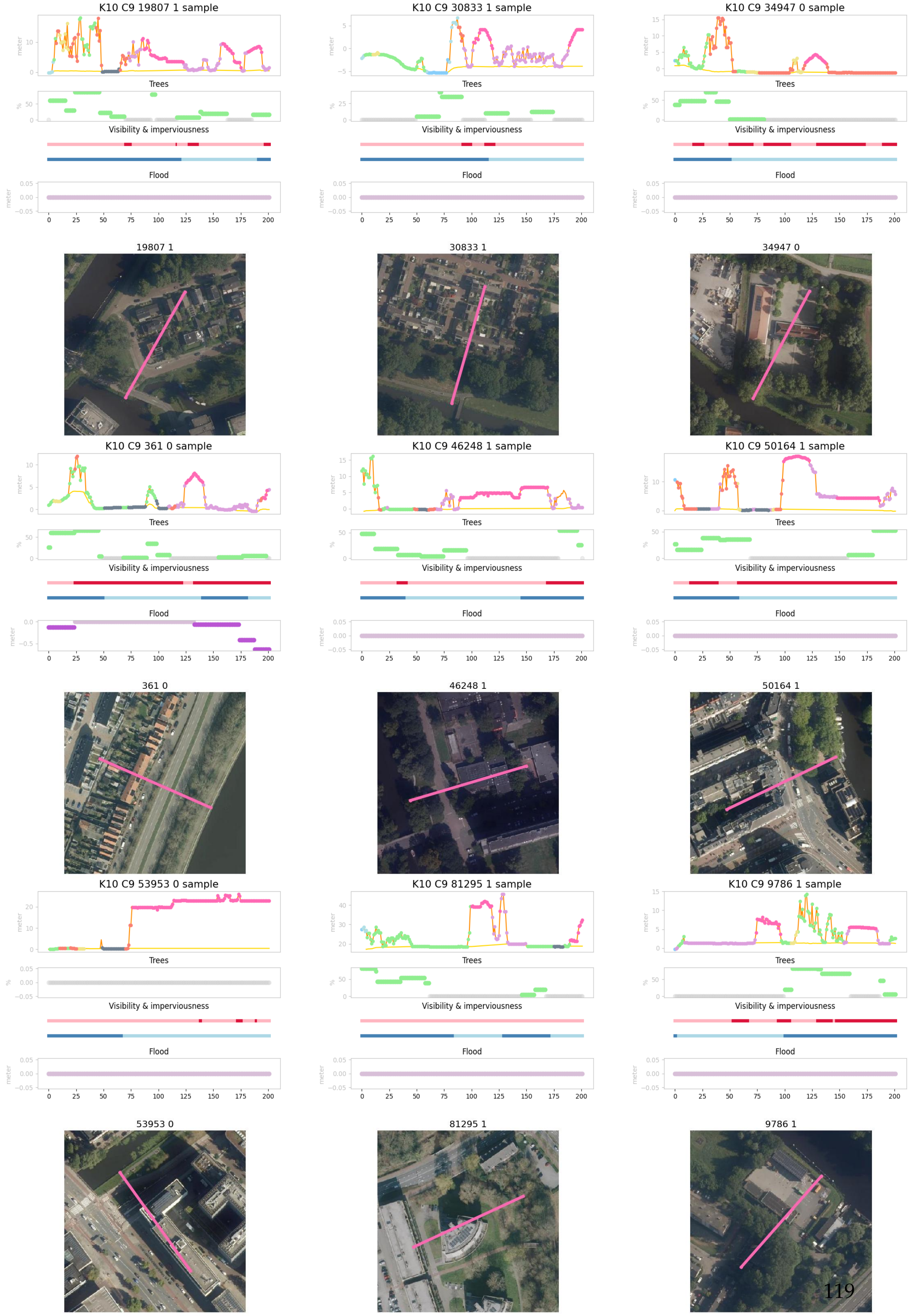


Figure E.11.: Sampled segments for cluster 9



F. Typology of urban river spaces in the Netherlands

This appendix shows the final product of the developed typology, identifying 6 types with 3 types containing subtypes. Each type is presented on a single page using the most representative segment and mean metric values, offering a quick overview of the results of this thesis.

Typology of urban river spaces in the Netherlands

The following pages summarize the types of urban river spaces. Values are given as mean values, and percentages imply ratio of segment covered by the metric. The spider plot shows the average values over the whole dataset as a dotted circle, and the standard deviation in light pink.

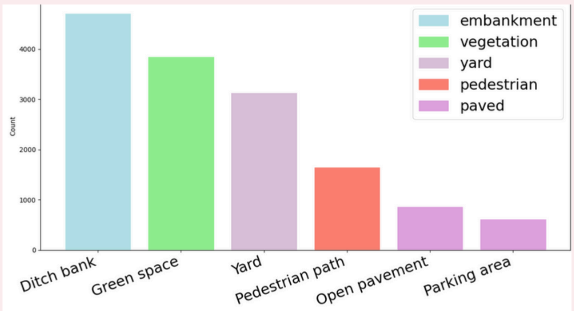
Building cover	*Tree %*	*Vegetation cover*	
High	Low	Low	**Type 1: Urbanized banks** 1A: Mixed hardscape greenway 1B: Open-edge urban river space 1C: Dense built-up river space
Low	Low	High	**Type 2: Lush natural floodplains**
Low	Low	Low	**Type 3: Barren paved land**
Low	Medium	High	**Type 4: Residential tree-covered river space** 4A: Visible areas 4B: Sloped areas
Medium	Medium	High	**Type 5: Balanced riverside corridors** 5A: Buildings and yards 5B: Set-back buildings and green riverbanks
Low	Medium	Medium	**Type 6: Mixed green strips**

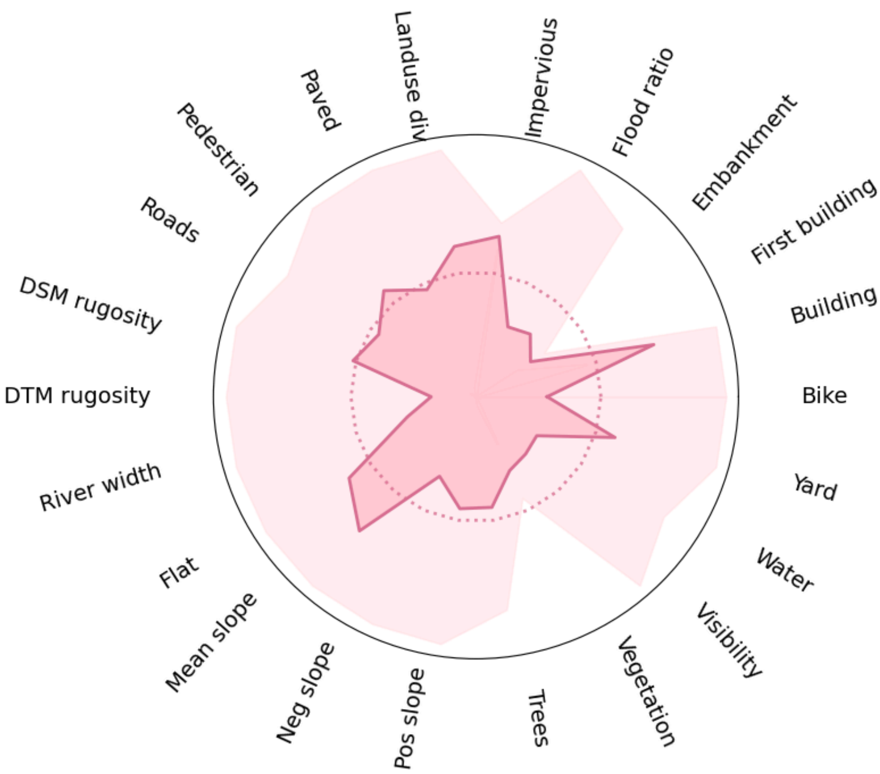
Urbanized banks
A: Mixed hardscape greenway

A building-covered residential river space with low vegetation levels but with more trees present than other subtypes for Urbanized banks. Green space is primarily present right next to the river.

33.9 % **37.0%** **8.7%**

First landuse distribution





Delineation

Flood ratio
14.5%

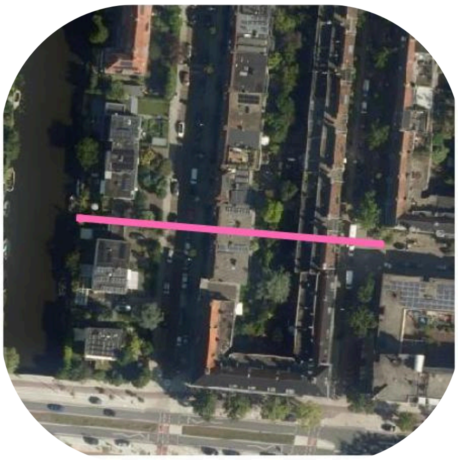
Visible space
67.7 %

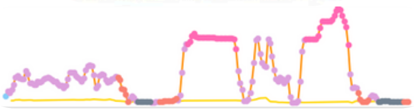
Distance to first building
18.9 m

Urbanized banks

B: Open-edge urban river space

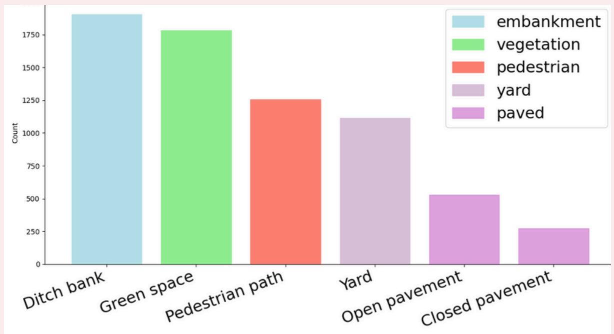
With buildings set back, this building-covered river space is more open towards the river, which allows bike paths to often be present in this space. Vegetation and tree presence is minimal, and the space tends to contain more open paved surfaces.

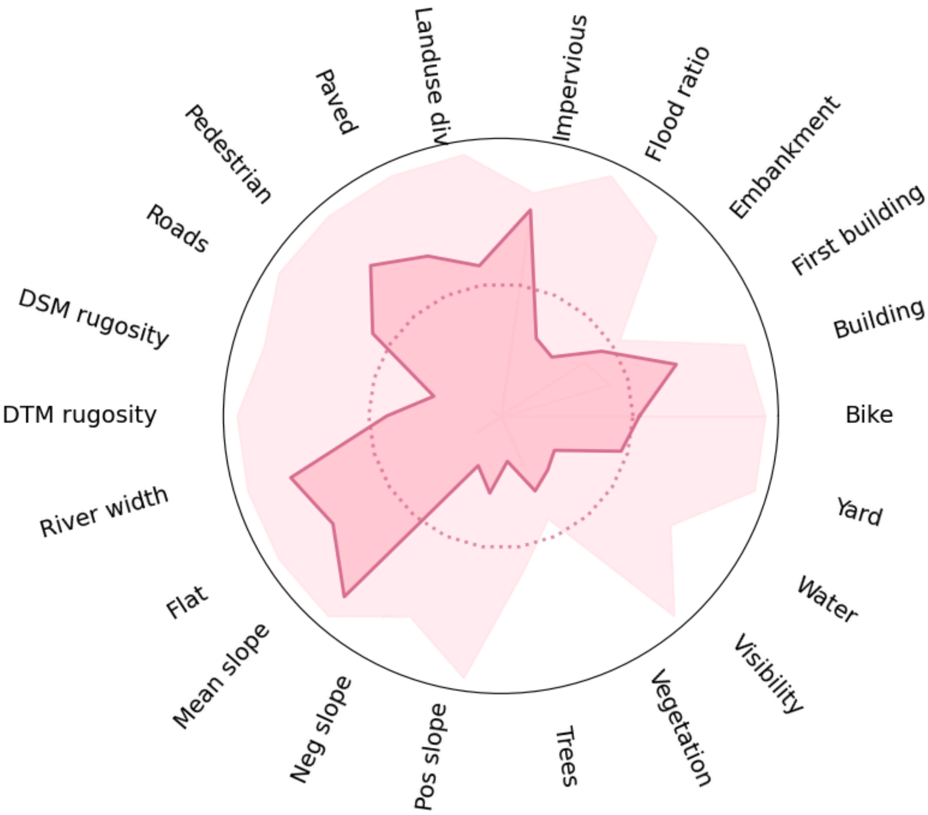




31.0% **24.6%** **7.7%**

First landuse distribution





Delineation

Flood ratio
14.9%

Visible space
67.5 %

Distance to first building
45.8 m

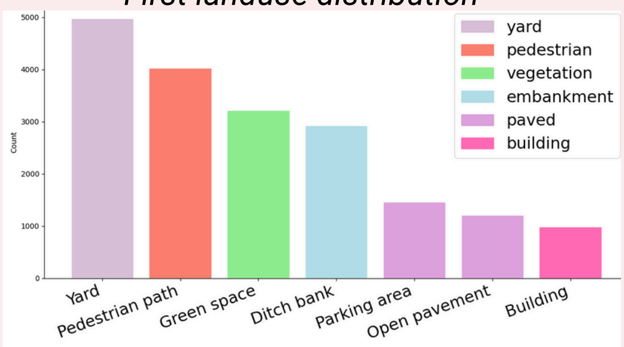
Urbanized banks

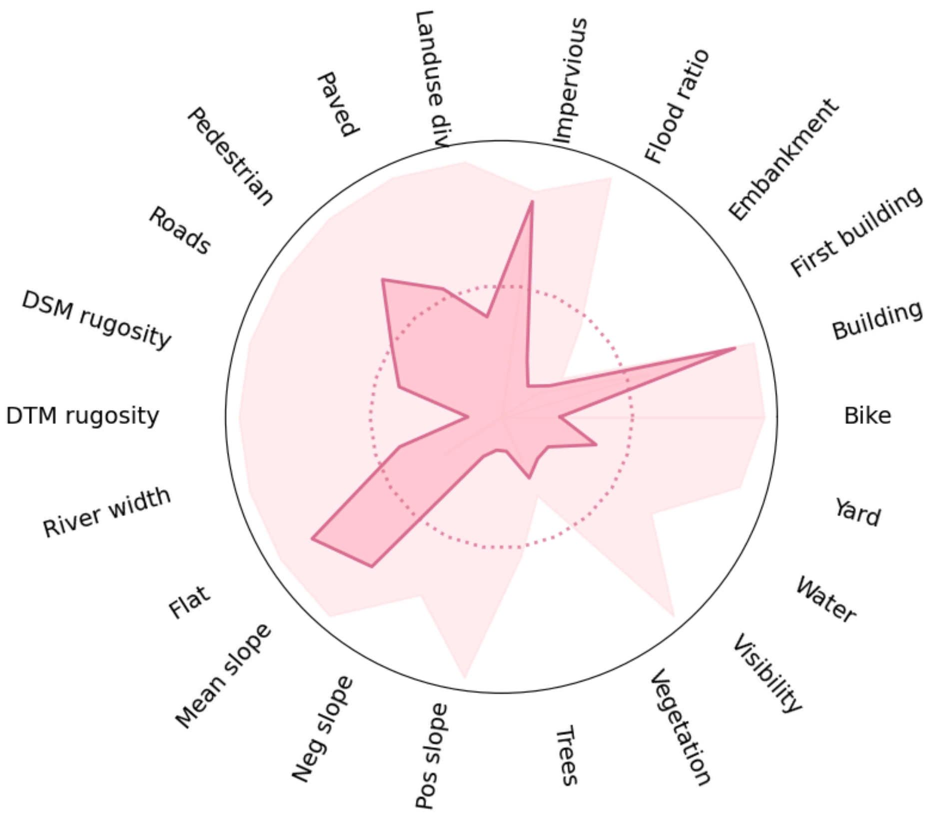
C: Dense built-up river space

A space almost completely covered in buildings, limiting the visible space, and offering very limited room for vegetation and trees.

44.9 % **22.8%** **3.3%**

First landuse distribution





Delineation

Flood ratio
12.7%

Visible space
67.0%

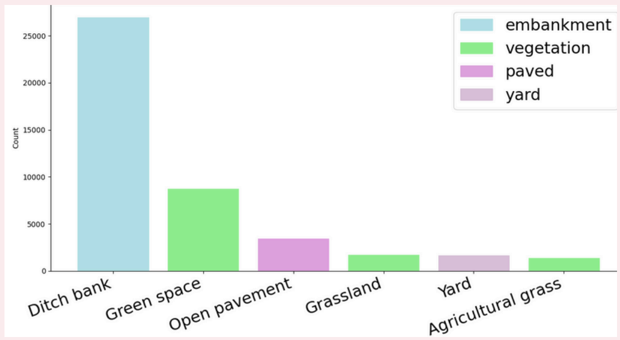
Distance to first building
13.1 m

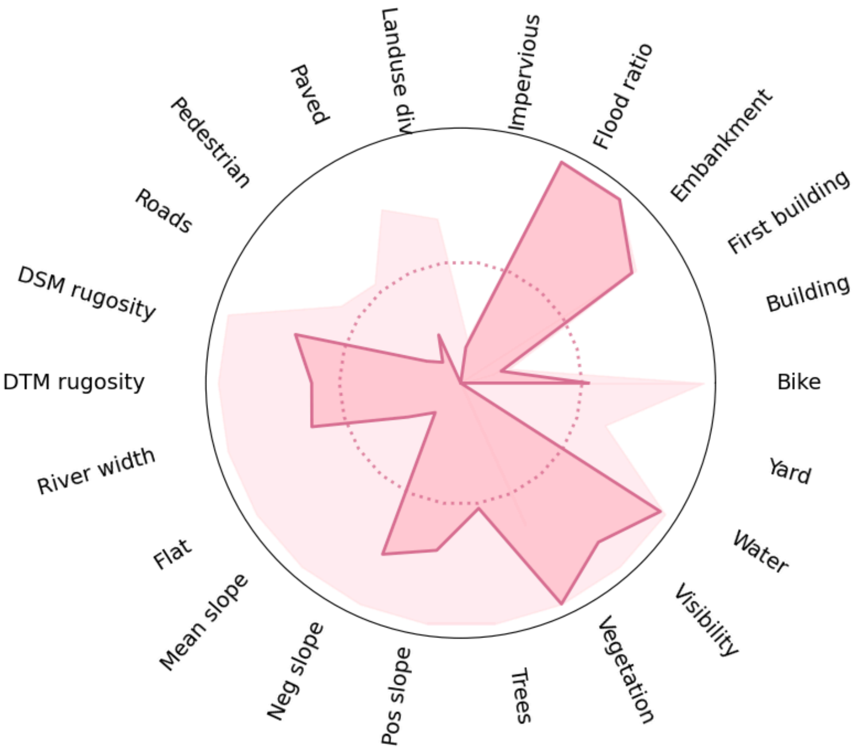
Lush natural floodplains

A green, often floodable, river space absent of buildings or has limited urbanized areas present. The green area is mostly covered by low vegetation instead of trees, and therefore the river is often visible wherever you are in the space.

0.4 % **40.4 %** **69.7 %**

First landuse distribution





Delineation

Flood ratio
35.3 %

Visible space
72.3 %

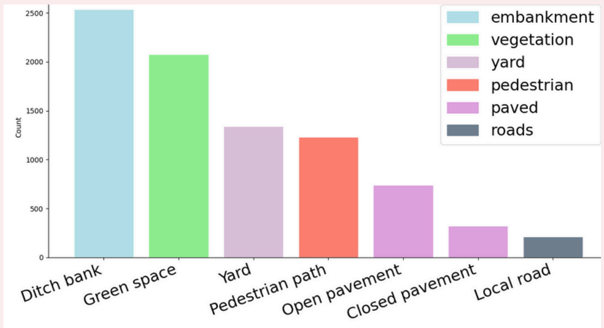
Distance to first building
99.5 m

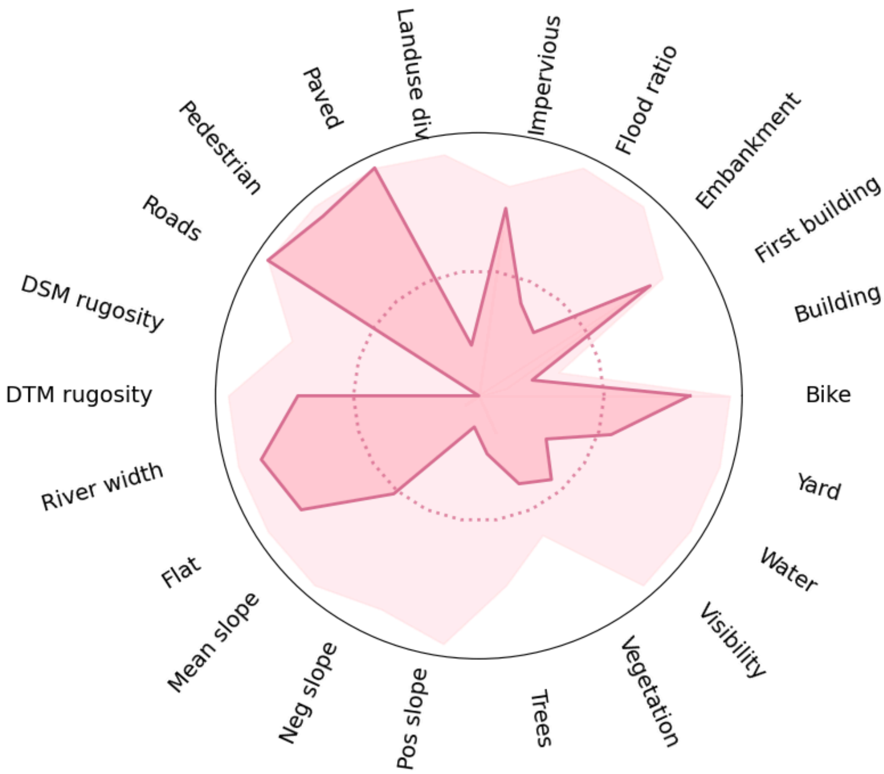
Barren paved land

A grey, flat river space, with limited vegetation and buildings. The space is often covered by pavement, roads, pedestrian areas, yard or bike paths and therefore has a high level of impervious surfaces.

3.3 % **27.4 %** **13.4%**

First landuse distribution





Delineation

Flood ratio
17.1 %

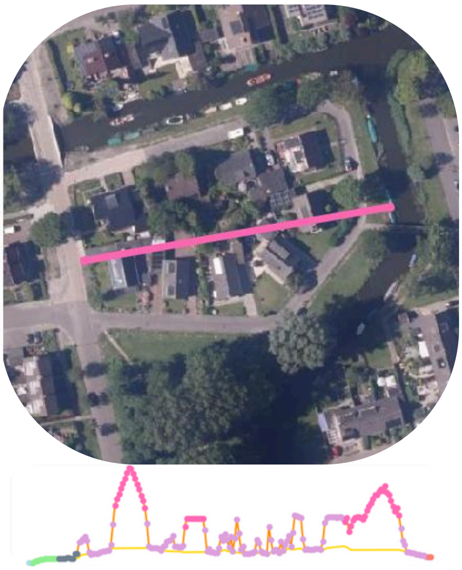
Visible space
68.9 %

Distance to first building
95.8 m

Residential tree-covered river space

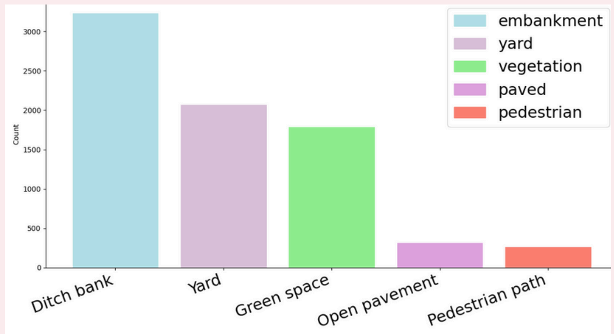
A: Visible areas

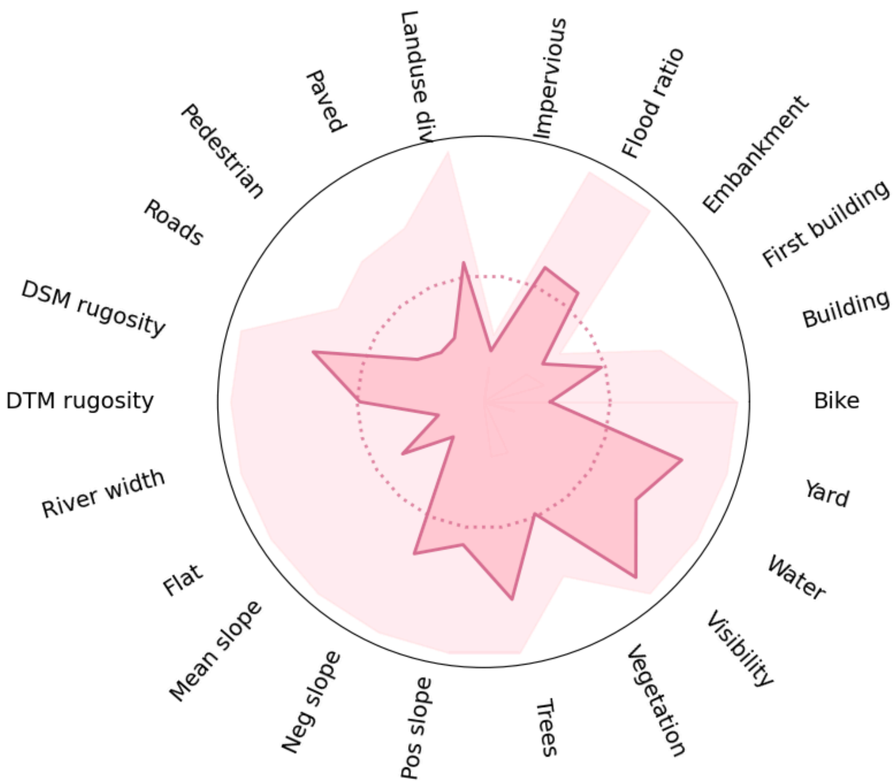
A tree covered river space with lower building cover, consisting of a mix of flat and sloped terrain. The space consists mostly of pervious surfaces. Most of the area in this type is considered visible space.



18.8 % **52.5%** **21.1%**

First landuse distribution





Delineation

Flood ratio
21.6%

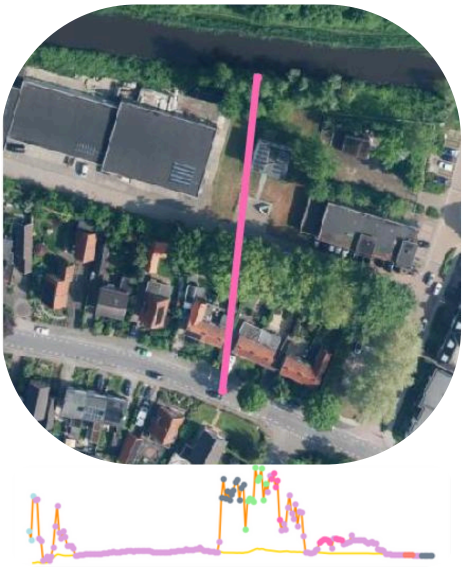
Visible space
72.8%

Distance to first building
21.3 m

Residential tree-covered river space

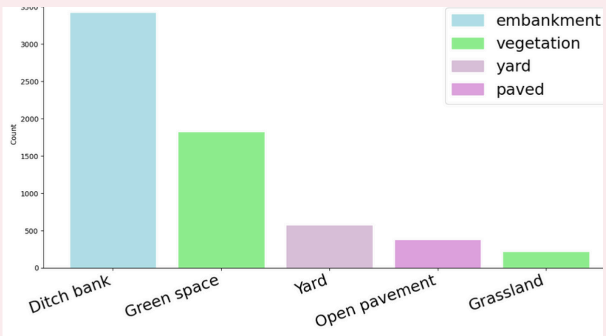
B: Sloped areas

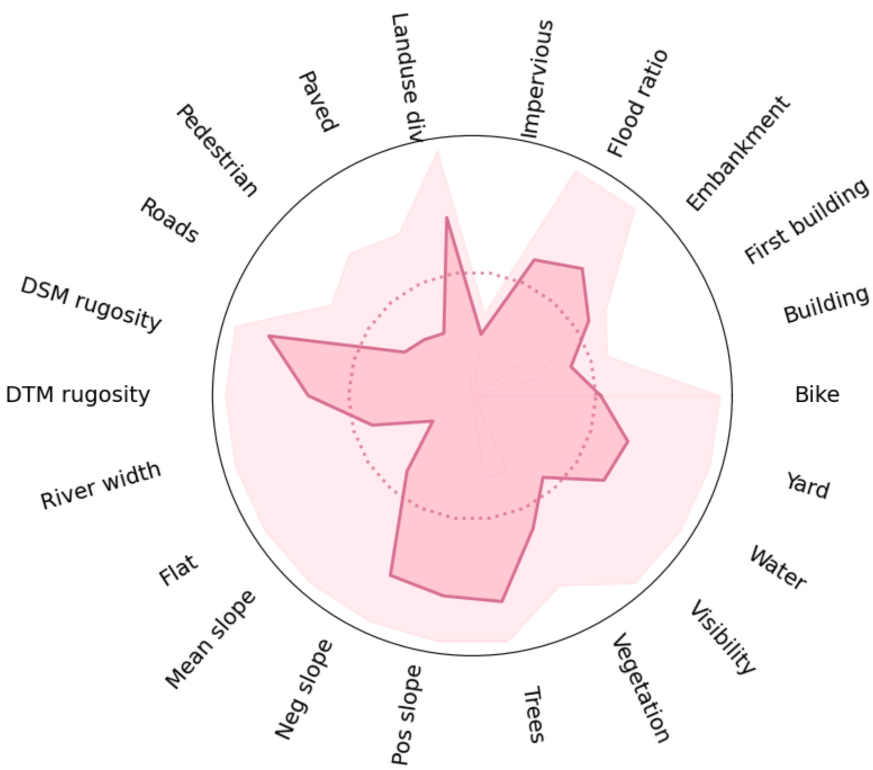
A tree covered river space with low building cover, characterized by the presence of sloped areas. The space consists mostly of pervious surfaces.



14.7 % **54.9%** **29.2%**

First landuse distribution





Delineation

Flood ratio
22.1%

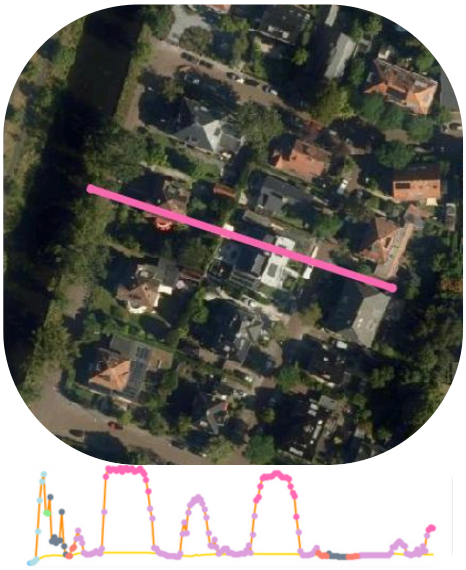
Visible space
68.8%

Distance to first building
60.6 m

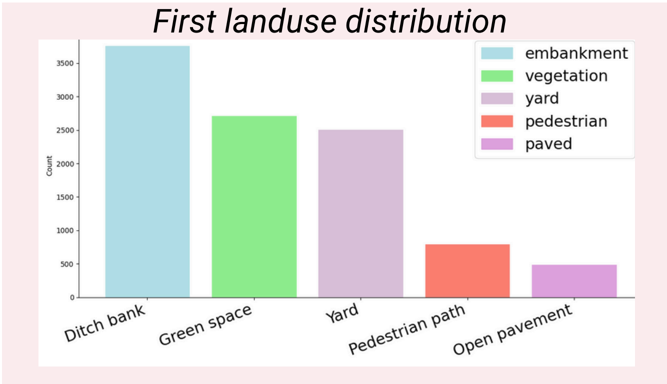
Balanced riverside corridors

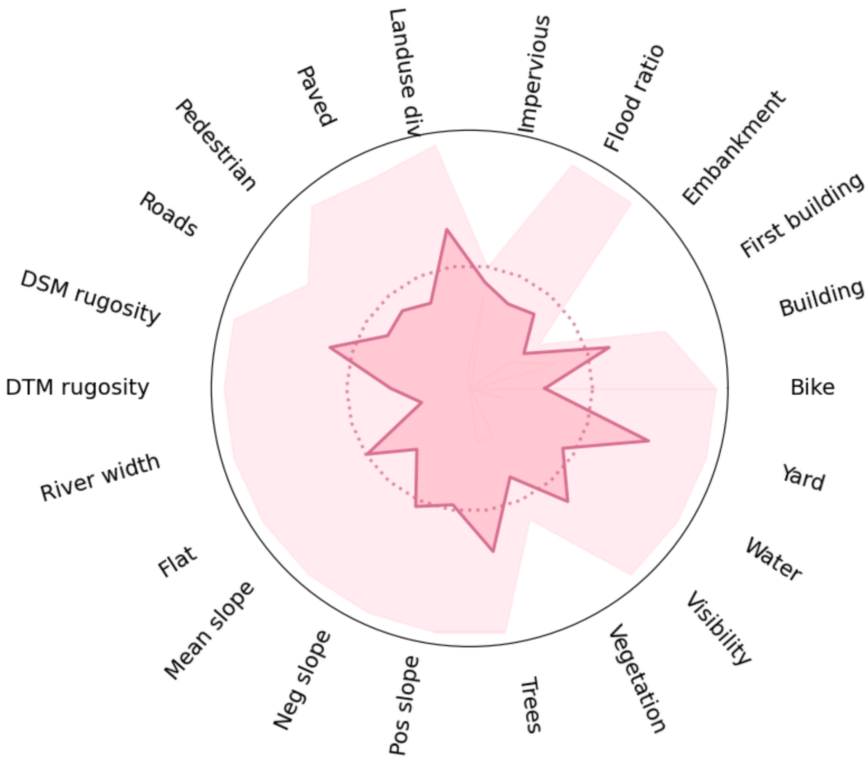
A: Buildings and yards

river space containing a moderate amount of buildings and their associated yards, often near the river.



25.0 % **47.2%** **14.2%**





Delineation

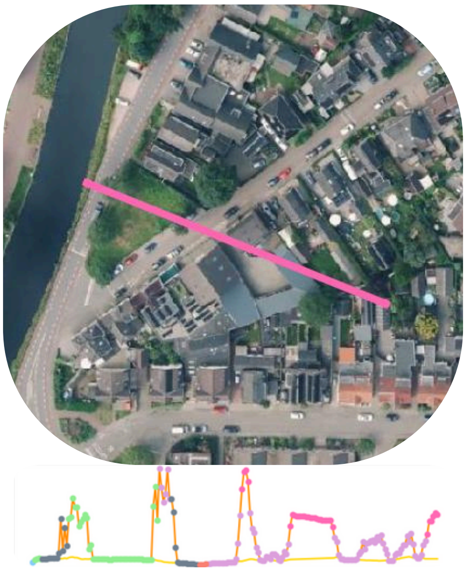
Flood ratio
16.3%

Visible space
70.2%

Distance to first building
19.2 m

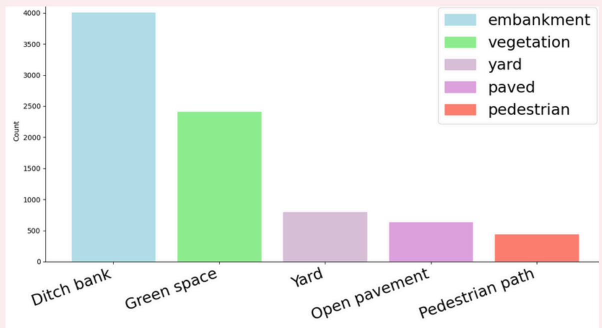
Balanced riverside corridors
B: Set-back buildings and green riverbanks

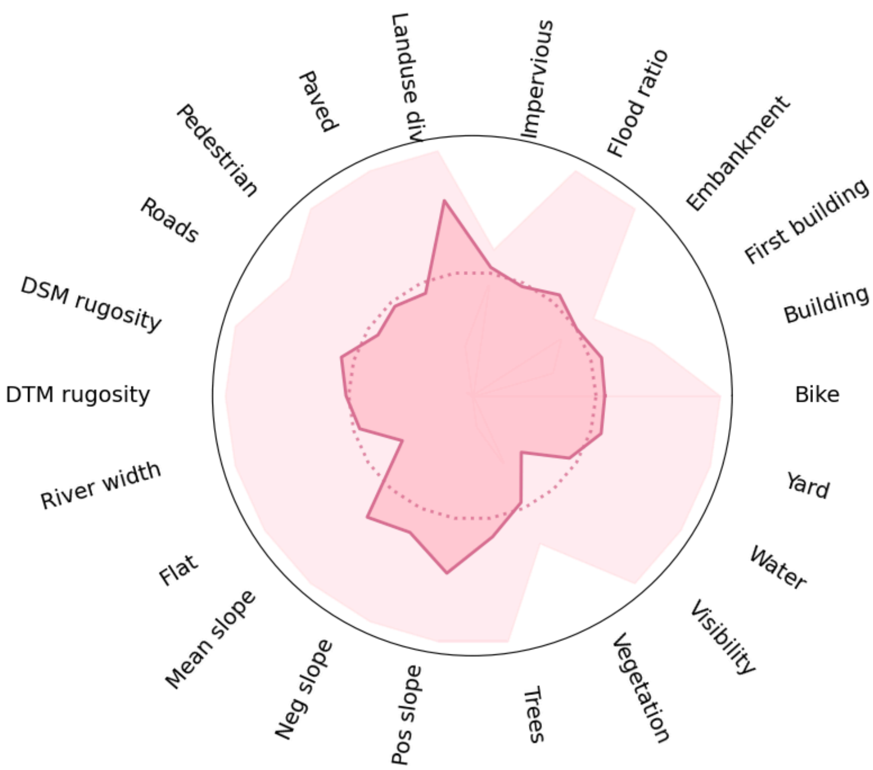
A space with a moderate amount of buildings, positioned further away from the river, often allowing a green area next to the river covered by vegetation or trees.



22.3 % **42.9%** **20.2%**

First landuse distribution





Delineation

Flood ratio
19.0%

Visible space
67.7%

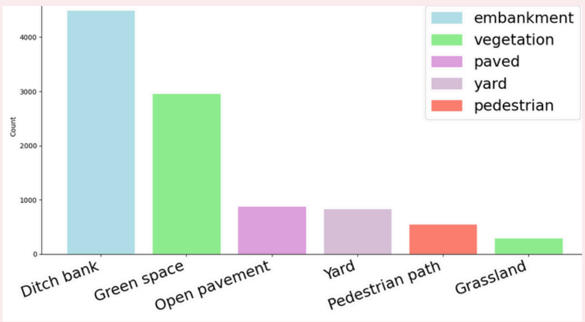
Distance to first building
52.4 m

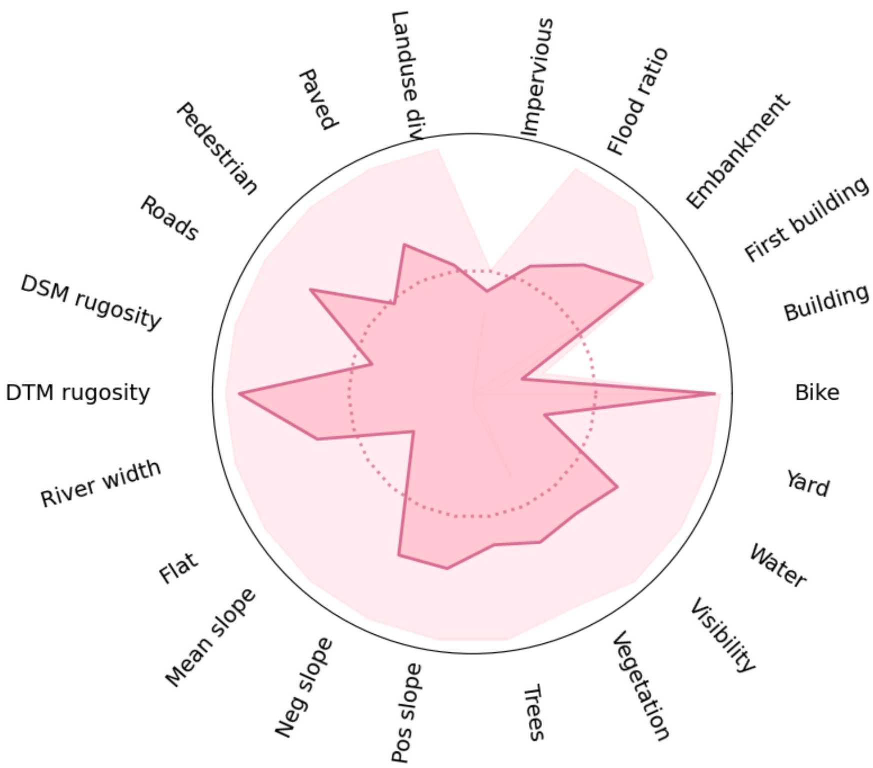
Mixed green strips

A vegetated river space laying in the built environment, often partly sloped, characterized by the extremely limited amount of buildings in the section, and the balance between vegetated and paved areas.

2.5 % **44.7%** **34.6 %**

First landuse distribution





Delineation

Flood ratio
21.2 %

Visible space
70.5%

Distance to first building
96.6 m

Bibliography

Apriliani, D. and Dewi, O. C. (2020). A study of cisadane riverside on riverbank development towards urban sustainability. *IOP Conference Series: Earth and Environmental Science*, 402(1):012011. Publisher: IOP Publishing.

Berghauser Pont, M., Stavroulaki, G., Bobkova, E., Gil, J., Marcus, L., Olsson, J., Sun, K., Serra, M., Hausleitner, B., Dhanani, A., and Legeby, A. (2019). The spatial distribution and frequency of street, plot and building types across five European cities. *Environment and Planning B: Urban Analytics and City Science*, 46(7):1226–1242.

Bisjak, A. (2005). Classification of the Aesthetic Value of Urban Rivers. *2005*.

Boongoen, T. and Iam-On, N. (2018). Cluster ensembles: A survey of approaches with recent extensions and applications. *Computer Science Review*, 28:1–25.

Brown, C. (2020). Urban rivers and their ecology. In *The Routledge Handbook of Urban Ecology*. Routledge, 2 edition. Num Pages: 11.

Burian, J., Sorbiová, K., Tucek, P., and Tučková, M. (2012). Possibilities of Delimitation of City Centers Using GIS. *International journal of humanities and social sciences*.

Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27. Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/03610927408827101.

City of Munich (2016). Isar-Plan – Water management plan and restoration of the Isar river, Munich (Germany).

Commons, W. (2025). File:Imagen de los canales concéntricos en Ámsterdam.png — Wikimedia Commons, the free media repository.

Djokić, V. (2010). MoRPHologY ANd tYPologY AS A UNiQUe diSCoURSe of ReSeARCH.

Dosary, A. A., Garba, S., and Mohammed, R. (2002). A Mechanism for Improving Service Delivery through the Accurate Delineation of Urban Boundaries.

Duranton, G. (2021). Classifying locations and delineating space: An introduction. *Journal of Urban Economics*, 125:103353.

Durán Vian, F., Pons Izquierdo, J. J., and Serrano Martínez, M. (2021). River-city recreational interaction: A classification of urban riverfront parks and walks. *Urban Forestry & Urban Greening*, 59:127042.

Fleischmann, M. (2023). Clustergram: Visualization and diagnostics for cluster analysis. *Journal of Open Source Software*, 8(89):5240.

Fleischmann, M., Feliciotti, A., and Kerr, W. (2022). Evolution of Urban Patterns: Urban Morphology as an Open Reproducible Data Science. *Geographical Analysis*, 54(3):536–558. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gean.12302.

Forgaci, C. (2018). Integrated Urban River Corridors: Spatial design for social-ecological resilience in Bucharest and beyond. *A+BE | Architecture and the Built Environment*, (31):1–382. Number: 31.

Gao, C., Dwyer, D., Zhu, Y., Smith, C., Du, L., Filia, K., Bayer, J., Menssink, J., Wang, T., Bergmeir, C., Wood, S., and Cotton, S. (2023). An overview of clustering methods with guidelines for application in mental health research. *Psychiatry Research*, 327:115265.

Gil, J., Beirão, J., Montenegro, N., and Duarte, J. (2012). On the discovery of urban typologies: Data mining the many dimensions of urban form. *Urban Morphology*, 16:27–40.

Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., and Briggs, J. M. (2008). Global change and the ecology of cities. *Science (New York, N.Y.)*, 319(5864):756–760.

Gurnell, A., Lee, M., and Souch, C. (2007). Urban Rivers: Hydrology, Geomorphology, Ecology and Opportunities for Change. *Geography Compass*, 1(5):1118–1137.

Hagberg, A., Swart, P., and Chult, D. (2008). Exploring Network Structure, Dynamics, and Function Using NetworkX. In *Proceedings of the 7th Python in Science Conference*.

Kondolf, G. M. and Pinto, P. J. (2017). The social connectivity of urban rivers. *Geomorphology*, 277:182–196.

May, R. (2006). "Connectivity" in urban rivers: Conflict and convergence between ecology and design. *Technology in Society*, 28(4):477–488.

Moretti, M. (2008). Cities on Water and Waterfront Regeneration: A Strategic Challenge for the Future.

Pattacini, L. (2021). Urban Design and Rivers: A Critical Review of Theories Devising Planning and Design Concepts to Define Riverside Urbanity. *Sustainability*, 13(13):7039.

Petrtýlová, R. and Matej, J. (2022). Waterfront character areas as the key elements in building the city - river relationship: Bratislava, Slovakia case study. *International Review for Spatial Planning and Sustainable Development*, 10(2):19–37.

Prominski, M., Stokman, A., Stimberg, D., Voermanek, H., Zeller, S., Bajc, K., and Zheng, N. (2023). *River.Space.Design: Planning Strategies, Methods and Projects for Urban Rivers Third and Enlarged Edition*. Birkhäuser.

Pugliese, F., Caroppi, G., Zingraff-Hamed, A., Lupp, G., and Gerundo, C. (2022). Assessment of NBSs effectiveness for flood risk management: The Isar River case study. *Journal of Water Supply: Research and Technology-Aqua*, 71(1):42–61.

Rahman, B. and Kautsary, J. (2024). River parks typology: A case study of river parks in Kalimantan. *IOP Conference Series: Earth and Environmental Science*, 1321(1):012017. Publisher: IOP Publishing.

Richardson, M. and Soloviev, M. (2021). The Urban River Syndrome: Achieving Sustainability Against a Backdrop of Accelerating Change. *International Journal of Environmental Research and Public Health*, 18(12):6406.

Rini, E. F., Rahayu, P., and Sinniah, G. K. (2020). Sustainable river management: Land use, building coverage, and infrastructure typology of the riverbanks. *IOP Conference Series: Earth and Environmental Science*, 447(1):012052. Publisher: IOP Publishing.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Schonlau, M. (2002). The Clustergram: A Graph for Visualizing Hierarchical and Nonhierarchical Cluster Analyses. *The Stata Journal*, 2(4):391–402. Publisher: SAGE Publications.

Silva, J. B., Saraiva, G., Ramos, I. L., Monteiro, F., Nunes da Silva, N., Câmara, C., Bernardo, F., Tourbier, J. T., and Westmacott, R. (2004). Classification of the aesthetic value of the selected urban rivers. Methodology. *Urban river basin enhancement methods-FP5 Project (2002-2005)*.

Sopena Porta, M. P. and Pellicer, F. (2024). Urban Planning and Landscape Projects on Urban Riverbanks in Europe: Comparative Study of the Ebro River, Zaragoza, and the Isar River, Munich. *Urban Science*, 8(4).

Wang, J., Fleischmann, M., Venerandi, A., Romice, O., Kuffer, M., and Porta, S. (2023). EO + Morphometrics: Understanding cities through urban morphology at large scale. *Landscape and Urban Planning*, 233:104691.

Wegmann, M., Zipperling, D., Hillenbrand, J., and Fleischer, J. (2021). A review of systematic selection of clustering algorithms and their evaluation. *ArXiv*.

Zingraff-Hamed, A., Serra-Llobet, A., and Kondolf, G. M. (2022). The Social, Economic, and Ecological Drivers of Planning and Management of Urban River Parks. In *Frontiers in Sustainable Cities*, volume 4, page 907044. ISSN: 2624-9634 Journal Abbreviation: Front. Sustain. Cities.

Colophon

This document was typeset using LaTeX, using the KOMA-Script class `scrbook`. The main font is Palatino.