Geographical Information Management and Applications

The added value of direct point cloud analysis in hydrology: A new method to derive streams from LiDAR data

Master Thesis

Stijn Ticheloven

Supervisors: Edward Verbree & Hans van der Kwast

Responsible professor: Peter van Oosterom

February 26, 2021

Universiteit Utrecht
TU Delft
University of Twente.
Wageningen UR

Preface

The writing of this Master's thesis gave me the opportunity to utilise all the academic skills I obtained during the past years. During the process, I became very motivated to get to the bottom of this challenging topic. Sometimes was required to go the extra mile to effectuate the satisfying results that can be found in this thesis. This hard work paid off in a thesis that I am very proud to present to you here.

I would like to thank Edward Verbree and Hans van der Kwast for their supervision. It was not possible to meet physically because of the Covid-19 pandemic. Despite these circumstances, they were always very involved in the process. Specifically, the specialistic hydrological knowledge of Hans was often very helpful and I am very grateful that he was willing to be a part of this research. Also, Peter van Oosterom has provided multiple times very profound feedback. His critical questions helped me to come to the interesting results in this thesis. Lastly, I would like to thank Raymond Nijssen for helping me review the Python-code.

I am proud to present this report to you. I consider this as the ultimate result of years of education. I am happy that I was able to contribute to the current hydrological scientific community, showed the added value of direct point cloud analysis and suggested an accessible, simple and accurate methodology for this. Also, I think this research highlights the potential of direct point cloud analysis. I believe that it would be very nice to implement the many future research suggestions.

Stijn Ticheloven

February 26th, 2021

Abstract

This research aims to investigate a new method for stream delineation with point cloud data in areas with low topographical relief, by using high-quality LiDAR point cloud data. Stream delineation is the derivation of streams, based on elevation data. By optimally using the quality LiDAR point cloud data, it is expected that better results in this field of study can be retrieved. The main objective of this research was to find and provide a proof of concept of an innovative method for stream delineation with direct point cloud analysis. Current methods predominantly use raster-based methods. Because of necessary interpolations that must be done during the creation of a raster terrain model, data is generalized and not optimally utilized for the analysis. The most commonly used conventional method is the D8-method. This method has limited direction options and, therefore, gives length errors. The new method handles this problem, as it can flow into an infinite number of directions. To identify the most suitable method that uses point cloud data more directly and to overcome the problem of the limited directions, an extensive theoretical review was performed and several prototypes were developed. The new method was selected by regarding three main requirements: accessibility, accuracy and simplicity. This means, respectively that the new method should be openly available for everyone, is more accurate than conventional approaches and is relatively easy to apply. This resulted in a new method: the absolute point-method. To ensure open accessibility of this method, the source code of this method can be found on: https://github.com/stn228/Absolute_point_method_Stream_delineation. This method looks for the lowest point within a specified search radius. To avoid getting in a local depression, an extendable neighbourhood and a maximum allowed course change was included in the algorithm. To examine the added value of the newly designed method, the results were compared to the results of a conventional stream delineation method: the deterministic 8 directions (D8), raster-based method. This research not only shows the added value of the newly designed absolute point-method but does also underline the great potential of point cloud data for hydrological analysis. The Root Mean Square Error of the D8 and the absolute point-method results were compared to a reference stream and the absolute point-method showed a significantly higher accuracy. Although this approach is very local, the absolute point-method offers a good alternative if conventional methods do not return satisfying results. It was demonstrated that, mostly in areas with very low topographical relief, the accuracy of this method is significantly higher. This shows the added value of the use of direct point cloud data for hydrological analysis. As hydrological applications become more and more important for proper water contamination management, this research advocates for a broader development of direct point cloud based hydrological tools. Finally, this research provides different suggestions for future research to scale this method. By adopting more efficient data structures and analysis methods for the point cloud, more advanced hydrological methods can become available.

Contents

List of figures

List of tables

Abbreviations

AHN	Actueel Hoogtebestand Nederland
D8	Deterministic 8 directions
DDT	Data-dependent triangulation
DEM	Digital Elevation Model
DSM	Digital Surface Model
DTM	Digital Terrain Model
GDAL	Geospatial Data Abstraction Library
IDW	Inverse Distance Weighting
LAS	LASer (file format)
LAZ	LAsZip (compressed LAS)
LiDAR	Light Detection And Ranging
MFD	Multiple flow directions
PDAL	Point Data Abstraction Library
RaDAR	Radio Detection And Ranging
RMSE	Root Mean Square Error
SAGA	System for Automated Geoscientific Analyses
SRTM	Shuttle Radar Topography Mission
TIN	Triangular Irregular Network

1 Introduction

Throughout the past decennia, possibilities for advanced digital mapping have become more widely available. New technologies have introduced continuously increasing computing power. This resulted in a rapidly increasing variety of data types and possibilities for analysis of the data. These emerging methods and techniques also resulted in new tools for hydrological analysis. Digital Elevation Models (DEMs) are commonly used for hydrological analysis (Longley et al., 2015). These gridded datasets represent parts of the real world with squared cells with a value assigned, that represent, for instance, elevation (Arun, 2013). Because of the wide availability and accessibility of gridded DEMs, the high information density and relatively low required computational power, this type of data seems very suitable to carry out hydrological analysis (Arun, 2013). Many raster DEMs are created by interpolating point cloud data, retrieved by using Light Detection And Ranging (LiDAR) scanning techniques. The raster-based approach is generally considered as accurate and the ease of this type of analysis is considered as an advantage. However, although raster techniques are widely used, Anderson and Ames (2011) suggest that information gets lost during interpolation of the raw point cloud data to a gridded format. With the emergence of the high-quality LiDAR point clouds, a solution to overcome this problem of information loss. Maintaining as much as possible data is required to retrieve accurate results for the analysis. Generally, the more data that can be utilised during the analysis, the higher the expected accuracy. A higher accuracy for stream delineation applications is mostly relevant for areas with low topographical relief. If direct analysis with the raw point cloud data can be executed effectively, significantly less loss of information will occur during the analysis because there is no interpolation needed. This increases the accuracy of the analysis (Anderson & Ames, 2011). Besides, raster-based approach are restricted to a finite number of directions to stream to. As a consequence, the model will show a cascading stream that follows the number of directions that it is restricted to; for instance, eight or sixteen directions (van Bemmelen et al., 1993). The conventional vector- and raster-based approaches do not seem to be the most optimal methods anymore. The relatively new point cloud representations possibly offer a solution, as they work without heavy interpolations and maintain the integrity of the data (Gabrisch, 2011).

1.1 Problem statement

Different researchers have identified the added value of point cloud-based approaches in hydrological analysis. Anderson and Ames (2011) explored a neighbourhood-based approach, while Gabrisch (2011) investigated this more direct point cloud analysis by using tessellations. This is considered a more direct approach as this method maintains the original elevation values of the points. Their approaches were proven and working on point cloud data. Also, the errors of the data are significantly lower than the gridded, conventional data (Gabrisch, 2011). Point cloud-based approaches are not new in hydrology. Different solutions with tessellations are found by researchers, but often data integrity is damaged constructing TIN-datasets by using gridded source data (de Azeredo Freitas et al., 2016; Rheinwalt et al., 2019; Zhang et al., 2018; Zhou & Chen, 2011). Gabrisch (2011) created a well-functioning algorithm based on tessellation to perform a catchment delineation and proved that direct point cloud analysis is more effective in areas with low topographical relief. However, their approaches are not easy to adapt and the created algorithms are highly complex and run on non-regular software.

Thereby, the performance in speed of the algorithms is lacking and the accessibility of these algorithms is insufficient. This research compares the different methods and creates an algorithm to delineate streams directly from point cloud data. Thus, this stream delineation tool derives streams efficiently from elevational data in an accessible manner.

Stream delineations model a part of the ecohydrological dynamics in a defined research area that specifically integrates the above-ground hydrological processes. Understanding and being able to predict these processes helps to regulate water quantity and quality within the researched area. The pressure on the human environment is increasing as overpopulation, urbanisation and climate change is becoming more and more significant. Effective sustainable policies are needed to maintain or restore the current water resources. To carry out an effective policy for water management, prediction of the behaviour of water in an area is very important. Mostly in underdeveloped areas, the need for water contamination and regulation is increasingly important. However, also in developed areas, it is important to expand knowledge to face the increasing severity of problems concerning water quality and quantity (IHE Delft Institute for Water Education, 2020). Because the above-described problems are occurring in various places, it is very relevant to be able to delineate streams for differing types of landscapes. Currently, accurate stream delineation, and thus creating an accurate ecohydrological model, is not well feasible in areas with low topographical relief. Concludingly, a stream delineation algorithm that is accurate in areas with low topographical relief, accessible, simple and one that performs well, concerning computing capacities, does not exist yet. This research aims to give an overview of existing methods and provides an accessible and accurate method to delineate streams more accurately in areas of low topographical relief. Therewith, it can be used as a tool to apply sustainable hydrological management and planning more effectively.

1.2 Research objectives

This research aims to delineate multiple streams in a predefined catchment area with low topographical relief using point cloud data. In this research, stream delineation is defined as the derivation of a water flow in a predefined catchment area, based on elevation models. This does not include the delineation of a catchment area, based on the derived streams. As stated above, Anderson and Ames (2011) and Gabrisch (2011) explored, among others, direct point cloud analysis. This research defines direct point cloud analysis as a method where the original z-values of the points in the raw point cloud are not modified. They successfully created algorithms that help to conduct direct point cloud analysis for hydrological applications. However, these algorithms do not focus on low relief areas or cannot yet be considered as ready to use tools. Because of its accessibility and simplicity, the conventional, grid-based approach is still dominant in hydrology. This research aims to create the fundaments for a ready-to-use stream delineation tool for using direct point cloud data. A comparison between different existing methods was made. The different approaches were being examined and an optimal method was identified and selected for further analysis. Open-source software was used, to enable future researches to get access to the tool. An advanced method to delineate stream in areas with low topographical relief is searched for.

This leads to the following research question:

"What is the best approach to delineate streams in areas with low topographical relief using LiDAR retrieved point cloud data?"

To answer the above standing research question accurately, the following sub-questions are drawn up:

1. "What different methods for stream delineation with LiDAR data exist and what strengths and weaknesses can be identified?"
2. "What approach with direct point cloud analysis is most suitable to delineate streams in an area with low topographical relief?"
3. "To what types of areas could a stream delineation with direct point cloud analysis be of added value?"
4. "Can an algorithm be created and applied in a suitable research area of the identified most suitable stream delineation method?"
5. "What is the added value of the selected approach in contrast to conventional methods to delineate streams?"

This research explores different existing approaches and selects the most suitable method. This selection is mainly based on three set requirements to meet the objectives of this research: accuracy, simplicity, accessibility of the algorithm. The accuracy should be higher than conventional existing methods in areas with low topographical relief; the algorithm should be able to handle big datasets and should be relatively simple to execute with normal computers. Lastly, the algorithm should be openly accessible and useable for everybody who wants to use it. This not only means that open-source software should be used, but also that an average hydrologist, without high-level programming skills, is able to execute the algorithm. The main objective of this research is to provide a proof of concept of a well-working stream delineation algorithm that meets these three requirements. If all these characteristics are sufficient, the algorithm accommodates the desired research objectives. Finally, this research identified the strengths and shortcomings of the designed method.

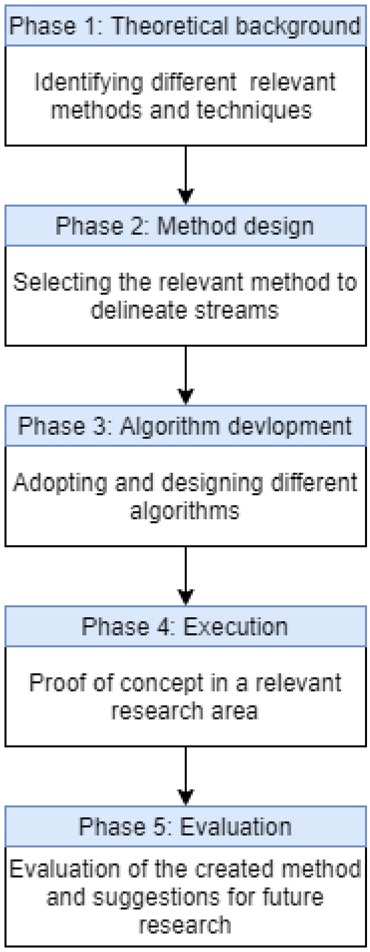


Figure 1: conceptual research approach

1.3 Research

To answer the main research question, a step-by-step approach was adopted. Four steps are identified, which must be executed consecutively. The first steps function as the fundament of later steps. These steps are graphically visualised in figure 1. In phase one, sub-question 1 was answered, phase 2 provides an answer on sub-question 2 and 3. Three candidate methods were identified. Based on the results of prototypes, the most suitable method was identified. Phase three answers sub-question 4 and phase four answers 5. Lastly, an evaluation of the created method was done. The results are reviewed extensively and suggestions for future research are done. This research gives an overview of different methods to prepare and process point cloud data for hydrological analysis. Based on the literature review, an algorithm was designed that meets the set objectives and requirements, mentioned in the previous paragraph. Finally, the algorithm was executed in two small sub-catchments as a proof of concept.

2 Theoretical background

Since the emergence of computer technology, stream modelling tools are increasingly frequent adopted. Different approaches were developed through the years. As this research aims to find an accessible and efficient tool to delineate streams, it is important to understand how these streams are extracted and what different approaches with LiDAR retrieved data already exist. This chapter elaborates on the different approaches that have been adopted in previous researches and gives an overview of the strengths and weaknesses that can be identified. This chapter lists the supposedly most useful methods to execute a stream delineation algorithm.

2.1 The concepts of stream delineation

The main objective of stream delineation is to model the flow of water in a specific area based on the topographical and geomorphological features of the surface. Therewith, a better understanding of complex ecohydrological systems can be achieved. This supports better water management as more customised policies can be made. Conventional, raster-based, algorithms do not require a selected starting or endpoint but define the streams in the research area, based on the continuous raster. Alternative models determine a flow path starting from any arbitrary point to its outlet (F. Zhang et al., 2018). However, it is also possible to delineate a stream from low to high. Then, the stream follows the steepest path up. This can be convenient if there are difficulties in finding the starting point. In many cases, the outlet point of a stream is easier to find, because this is the point where all the water flows to in the (sub-)catchment. However, the risk exists that a stream will be delineated away from the actual stream because the surface outside of the river bed is higher than points upstream. The stream can be delineated, based on the flow direction that can be extracted from an elevation model. The flow directions are based on the direction of the slopes between different measurement points. A stream network can be extracted, based on the determined directions of the flow (F. Zhang et al., 2018). This maps the drainage pattern in a specific area resulting in a comprehensive stream network existing of many different streams with varying significance. Raster-based approaches commonly order these streams, based on a flow accumulation or Strahler order, which is explained later in this chapter in section 2.4.1. Alternatively, the flow path is based on the steepest descent. The flow direction of the water naturally follows the path of steepest descent (F. Zhang et al., 2018). Furthermore, some important topographical characteristics of a stream must be taken into account when creating a stream delineation algorithm: a stream cannot cross itself, return to the path where it came from and an arbitrary maximum course change of the flow path should be considered (Anderson & Ames, 2011).

2.1.1 Conceptual model

To create and understand the new stream delineation method, it is important to understand the importance of the decisions that are made for creation. In all cases, the stream follows its way down or up over the modelled elevation surface. The main objective of stream delineation is to identify the direction of the flow at specified points on the surface. The results of the stream delineation are dependent on the quality and type of input data that is used. The stream flows over the surface in a specific direction that is dependent on the type of data.

From a specified starting point, the stream flows over a specified distance towards the next point and this process is repeated until the stream reaches the outlet. Mainly three methods can be distinguished and are discussed in this research. Traditional stream delineation with rasters stream from grid cell to grid cell and are restricted to the cardinal eight directions. With a triangular representation, the flow streams towards a connected neighbouring point; with individual points, the stream flows towards a point at an arbitrary distance. This process of identifying the path downward over a surface is iterated until it reaches the outlet of the research area. Mostly in areas with low topographical relief or with a highly complex surface, minor details can affect the final results of the stream delineation model. For example, errors occur when algorithms end up in a local depression (sinks). The influence of the data preparation phase and therewith the chosen input data type is an important aspect of this research and is, among other things, discussed in this chapter. A choice can be made for a highly accurate surface representation, which is more complex to work with, or for a more simplified representation of the earth's surface, which is more suitable for analysis but possibly impacts the accuracy of the results. These decisions strongly influence the final results of the analysis and therefore should be chosen with care. Based on the theoretical background and the created prototypes of the selected candidate methods, a choice was made for the algorithm that is developed. This research uses high-quality point cloud data to delineate a stream in a research area. Then, based on the theoretical review, the most suitable representation of the data was selected to perform a stream delineation that preserves the original high-quality point cloud data as much as possible. As stated before, when selecting the data representation and the stream delineation algorithm the accuracy, simplicity and accessibility of the model are decisive for the choice. This process is conceptually shown in figure 2. This chapter discusses the different possible methods. Methods that comply with these three requirements are identified in this research.

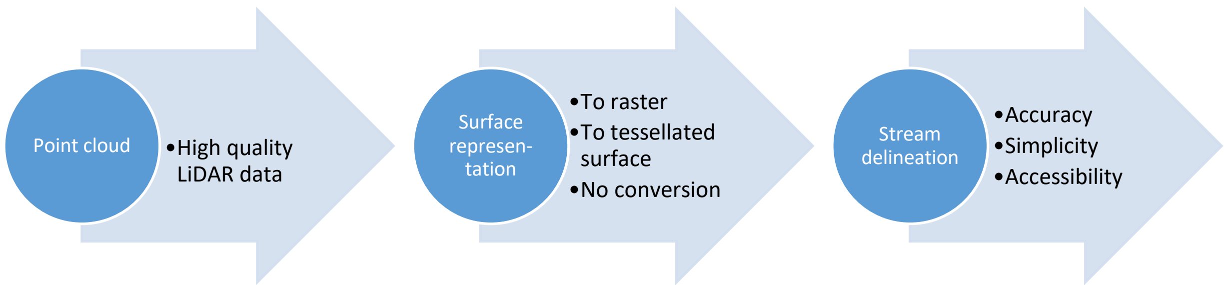


Figure 2: Conceptual model.

2.2 Analysis steps

Traditionally, stream delineation analyses are conducted with raster-based elevation models. It can be assumed that the more accurate the elevation is modelled, the better the results of the analysis are. Elevation data can be gathered and preprocessed in different manners. This research uses LiDAR data. Although LiDAR data can be considered highly accurate, different applications with the data can have a great impact on the results of the analysis. More specifically, the results of a stream delineation analysis can be influenced during mainly four stages:

- Data gathering;
- Data preparation;
- Preprocessing;
- Processing.

Preprocessing of the data by collectors are not included in these steps as this is commonly done by the data-gathering agencies. It is important to note that this potentially has an impact on the data. For example, the chosen methods during the scanning process impact the resulting point cloud dataset and thus, the results of the research. The data can be prepared to make it easier to work with. Commonly, point clouds are very big datasets with not many ready-to-use tools available. To make the data easier to process, the data sometimes is being converted to other data types such as gridded- or triangulated datasets. Furthermore, to minimise the occurrence of expected errors, the data can be preprocessed. Examples of this are terrain reconditioning algorithms, such as filling the voids and sinks in the data or river burning (Wang & Liu, 2006). Finally, the last step is the processing of the data. This is where the actual stream delineation algorithm is being executed. This chapter discusses the different methods and possibilities step-by-step.

2.2.1 Data gathering

There are different methods available to collect topographical data. Remote sensing techniques enable researchers to measure electromagnetic radiation coming from an object. The retrieved data can be translated into information about the ground object (Meesuk, 2017). Mainly two types of remote sensing can be distinguished: active and passive. Active remote sensing techniques are commonly used for measuring the height of land surfaces. Although there are passive remote sensing techniques, such as Structure from Motion and photogrammetry, to create point cloud elevation models, these techniques are less accurate than the more recently developed active scanning techniques (*GIM International*, 2020). Two techniques are available and can be generalised to Radio Detection and Ranging (Radar) and Light Detection and Ranging (LiDAR) applications (Clevers, 2020).

Radar

Radar retrieved topographical surface representation is the most common method to acquire elevation data. Active microwave sensors illuminate the earth surface and measure the backscattered signal. With the time delay and strength of the received signal, the distance to the target can be determined (Clevers, 2020). As a result of the radar sensing, a gridded Digital Elevation Model (DEM) can be created. Each pixel in the dataset depicts an elevation value. Global satellite programs, such as the Shuttle Radar Topography Mission (SRTM) of NASA, result in large-scale and cost-efficient DEMs. Because of the wide availability and sufficient quality and reliability, these DEMs are suitable for various hydrological analyses. However, the coarse resolution of global Radar DEMs limit the accuracy of stream models with this type of data as input. The finest resolution of the SRTM is a DEM with 30-meter pixels, while the LiDAR retrieved AHN DEM has a 0,5-meter resolution (AHN, 2019). Thus, these global radar DEMs are useful for regional and global scale studies to identify patterns that should be further examined using higher-resolution datasets (Acharya et al., 2018). Mostly because of the regional nature of hydrologic basin analysis, specifically stream delineation, a more specific elevation model should be adopted for more accurate and refined results. As stated before, more accurate elevation models can be retrieved with LiDAR technologies.

LiDAR

LiDAR is an active optical remote sensing technique that generates a representation of the earth's surface with point clouds. Aerial scanning is only discussed as this is the most suitable method to create an elevation model with LiDAR. With airborne LiDAR scanning, a laser scanner is attached to a plane and transmits laser beams to the earth surface. The beams are scattered back and then are analysed by the receivers in the sensor. Based on the calculated travel time of the laser beam, the distance to the earth, or a specific object on the earth, can be determined (Clevers, 2020). The main difference from Radar is that LiDAR uses much shorter wavelengths in the electromagnetic spectrum; the ultraviolet, the visible or the near-infrared (Clevers, 2020). This commonly results in a large dataset that consists of independent LiDAR-points with x, y and z values: a point cloud (Longley et al., 2015). These point clouds provide a 3D visualisation of the earth surface and can be considered as a valuable addition to the two traditional forms of earth modelling, vector and raster representations. The point-density of the point cloud and thus the accuracy is dependent on the quality of the measuring equipment and the flight planning settings. More concrete, this is dependent on the scanning angle, pulse repetition and scanning frequency for the sensor and the flying altitude, speed and overlap for the flight planning (Höfle & Rutzinger, 2011). One of the great advantages of LiDAR retrieved data is that the geomorphological specifications of a surface can be mapped with much detail and because of the active data acquisition process with laser scanning, the point cloud is even measured under vegetation cover (Höfle & Rutzinger, 2011).

Each point in the point cloud has its own x, y and z value. These point clouds provide a 3D visualisation of the earth surface and can be considered as a valuable addition to the two traditional forms of earth modelling, vector and raster representations. One of the great advantages of LiDAR retrieved data is that the geomorphological specifications of a surface can be mapped with much detail and because of the active data acquisition process with laser scanning, the point cloud is even measured under vegetation cover (Höfle & Rutzinger, 2011). The usage of LiDAR to create point cloud representations of the earth's surface is emerging, because of the high accuracy and the potentially wide range of applications of LiDAR in, for example, forest management, urban planning, hydrological analyses or flood control (Aguilar et al., 2010). In particular, point cloud representations are suitable for analyses where details possibly affect the results. This is because LiDAR depicts very detailed elevation information of a specific area (Höfle & Rutzinger, 2011). An example of this is in hydrological analyses in areas with low topographical relief. High accuracy of the actual elevation under vegetation is important to understand the natural stream of the water.

The gathered data is commonly preprocessed by the data collectors. The adjustments of the data should result in more accurate representations, but it is important to note that this could affect the data. A frequently used preprocessing step is filtering. Using a filtering algorithm, ground and non-ground points can be distinguished. Filtering on ground points, for example, enables researchers to conduct bare earth analyses. Also, outliers can be filtered out from the data. This requires intelligent algorithms with diligent parameter adjustment for such big datasets (Chen, 2007). Furthermore, tiling and compressing data could influence the data. This helps users to access the big datasets more easily, but the data are reorganised in this process and could affect the data, mostly at the edges of a tile (Chen, 2007).

Besides the various advantages of LiDAR, there are some downsides to the use of LiDAR. First, aerial LiDAR data collection is expensive because a plane, a pilot and a high-quality laser scanner is required to acquire the information. Thereby, the plane can scan a limited area at a time. This makes the gathering of the data a very time consuming and logistically complex process. Although the cost-effectiveness of LiDAR scanning is currently continuously improving by, for example, increasing the pulse frequency of a single scanner or by mounting multiple scanners on a plane, it is still an expensive process (Johnson et al., 2014). Furthermore, since point clouds are a relatively new data type, next to the traditional vector and raster data types, there are not many easily accessible applications for use of the data. Thereby, the resulting datasets are often very demanding of computer processing capabilities because of their great data density. Therefore, many researchers choose to modify the datasets to traditional data types before they use them. This results in (partial) loss of a previously given advantage: the high accuracy.

2.2.2 Data preparation

As stated before, the relatively high complexity of point cloud data could force researchers to modify the data to a more useable data format. The main goal here is to prepare the data for further analysis, without using too much accuracy of the origin data. The topographical modelling of a terrain is often associated with a gridded visualisation, known as the Digital Elevation Model (DEM). Accordingly, point clouds are frequently converted to these gridded DEMs to make them useable for the analysis phase. However, there are many different possibilities to create an accurate model or representation of the earth's surface. This results in terrain models in different dimensions, with different features and different characteristics (Gold, 2016). Logically, it can be stated that these different models influence the final results of the research. Generally, a researcher can choose to prepare the raw point cloud data in three different manners:

1. Conversion to raster datasets;
2. Conversion to tessellated datasets;
3. No conversions.

With a conversion to a continuous surface (raster or TIN), the stream direction of the flow can be retrieved for the whole area. No conversions presumably lead to a more local approach where the direction of the flow is retrieved per point. This paragraph discusses the different possible conversion methods relevant to this research. Later in this chapter, the different methodological approaches are discussed.

2.2.3 Conversion to raster dataset

Traditionally, rasterised data is broadly used for hydrological analysis. Before the emergence of LiDAR technology, elevation data was mostly gathered with Radar which results in rasterised datasets (Clevers, 2020). The pixels are a very efficient way to visualise the surface elevation with single values, in 2.5D (Höfle & Rutzinger, 2011). Thereby, as stated before, there are many applications available that require raster input for a wide variety of analyses. Therefore, point cloud data is regularly transformed into raster data. This can be achieved with calculations that assign values to the pixels in the grid, based on the measured x, y, z values assigned to the points in the point cloud. Different interpolation methods perform such calculations. Mainly two approaches for point cloud to raster interpolation can be identified: (1) deterministic methods and (2) geostatistical methods (Xiaoye Liu, 2008). For both, respectively one specific method has been lighted out as an example: (1) Inverse Distance Weighting (IDW) and (2) Kriging.

Deterministic methods

The deterministic approaches assume that each input point has a local influence that diminishes if the distance from the point increases (Xiaoye Liu, 2008). IDW assumes that points closer to a specific location have a stronger influence than the farther points. Within a window, an adjusted number of points are selected and a weight is assigned to it based on the proximity of the point. This is visualised in figure 3. The determination of the search window can be fixed, based on a radius, or variable, based on a number of points that must be included in the search window. IDW is particularly useful if the input data is dense and well-distributed (Xiaoye Liu, 2008). Moreover, because IDW contains original values, a proper approximate representative elevation can be created. In the case of interpolation of point cloud datasets, it can be assumed that the first criterion is met. However, LiDAR retrieved point clouds are known for their irregular distribution of points because the laser reflects randomly (Clevers, 2020). This potentially results in less accurate results of an interpolation of point cloud data. Thereby, IDW is a method that uses weighted averages and is not capable of considering survey points outside the search window. This could result in the ignorance of topographical characteristics, such as steep ridges. These shortcomings are of great influence in hydrological analyses and could lead to ignorance of important features of the surface. Therewith, the result of the model could be impacted drastically. Although these methods are usually relatively easy and accurate, deterministic methods are not always suitable for system modelling, because do not take the model of spatial processes within the data into account (Xiaoye Liu, 2008).

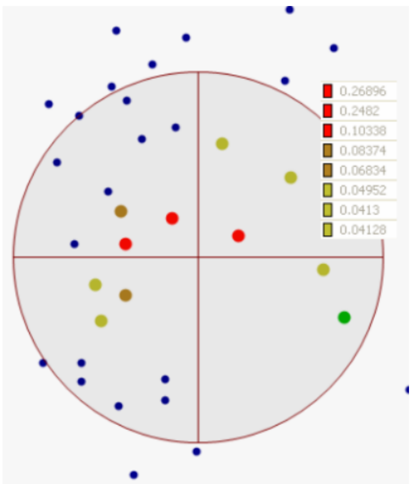


Figure 3: IDW Search Window where closer points have a higher assigned weight (Esri, 2020).

Geostatistical methods

The spatial correlation of the data is considered in geostatistical interpolations (Xiaoye Liu, 2008). Spatial correlation assumes that the closer a specific point is located to the sample point, the more it is related. As mentioned before, an example of a geostatistical method is Kriging. Therewith, it assumed that there is a spatial correlation between the sample points. The correlation can be used to explain the variation in a surface (Xiaoye Liu, 2008). In essence, it is a statistical approach that can be used, using the mathematical formula of kriging to predict the surfaces of an area. By creating variograms and covariance functions, the statistical dependence can be estimated. Therewith the unknown values in the surface can be filled in (Esri, 2020). Essentially, Kriging is a technique that weights averages of weighted distances between sample points and estimated locations and mutual distances among sample points (Xiaoye Liu, 2008). This results in more sophisticated, and therewith more complex, surface models. However, it assumes a specific pattern in the data while this is not always true for complex surfaces. A generalisation of the data is manifested. This generalisation is applied in all both geostatistical as deterministic interpolations. Hui et al. (2016) introduces a multilevel Kriging filtering algorithm with a high accuracy and low errors in comparison to other interpolation methods for LiDAR data. The combination of the relatively easy applicability and the high performance of geostatistical interpolations this a suitable choice in the data preparation phase. However, although Kriging, and other geostatistical interpolation methods, are considered as powerful and efficient tools, the transformation with interpolations to a gridded dataset results in data loss. This causes errors in the pixel values and the analysis is generally constrained by an eight-direction flow algorithm. This potentially impacts the final results of the model (Gabrisch, 2011).

Concludingly, it can be stated that a conversion to gridded data has well-developed methods, where the resulting datasets are DEMs of good quality. However, the accuracy of the resulting raster grid is limited and could impact the final results if the details in the earth surface matter for the results of the research, for example in areas with little differences in relief. Thereby, raster representations are 2.5-dimensional as they only contain height information and no volume (Garbrecht & Martz, 2000). Thereby, most conventional methods only use the x and y values to determine the weight of the specific cell. For instance, IDW bases the weight on distance. This weight is used for the calculation of the average z-value for the specific cell. This makes the gridded representations a constrained simplified image of the real world wherein complex geometries are not easy to depict. An approach where the interpolations are more dependent on the elevational data is desired for more immersion of the elevation model.

2.2.4 Conversion to tessellated datasets

The use of tessellated datasets as a representation of the earth's surface is often suggested as an alternative for raster solutions. Often, tessellations manifest themselves through triangulations. A prevailing triangulation type is the Triangulated Irregular Network (TINs). One of the great advantages of TINs is that the source data point-values are not modified. By creating a TIN, the survey points are converted to a continuous surface, existing of vertices that are connected with edges by triangular facets. If only the edges and vertices are used as a network for the analysis, the triangulated dataset is no longer a continuous representation of the surface area, as the faces of the triangles are not taken into account and do not represent anything. The raw data is more or less preserved in the interpolated surface as the vertices are the original measurement points (Gold, 2016). For triangulation of datasets, different methods can be identified. These different methods can be categorised by dimensionality: 2D, 2.5D and 3D; or: TINs, Data-Dependent Triangulations (DDTs) and tetrahedralizations.

Triangulated Irregular Networks (TINs)

Creating a TIN can be done with different approaches. As stated before, TIN representations are created from data where the data loss is minimal. Thereby, this type of interpolation of point cloud data can be considered as a more sophisticated techniques for terrain analysis that maintain the high information density of the LiDAR data (F. Zhang et al., 2018). Also, the TIN-structure enables to store the available data more efficiently and perform better in feature representation (Zhou & Chen, 2011). Furthermore, the processing speed is relatively good in these data structures, both TIN and Voronoi, because it is possible to create hierarchies in the data (Gold, 2016). One of the most commonly used calculation for TIN creation is the Delaunay triangulation. The Delauney triangulation fulfils the 'empty circle criterion'. This criterion is visualised in figure 4 on the next page (Verbree & van Oosterom, 2003). As a result, the high information density can be maintained, but they do not ensure that all hydrologic features are modelled as accurate as possible (Nelson et al., 1994). Lastly, the irregularly sampled point, characteristically of a LiDAR retrieved point cloud, is very suitable for creating triangulated networks.

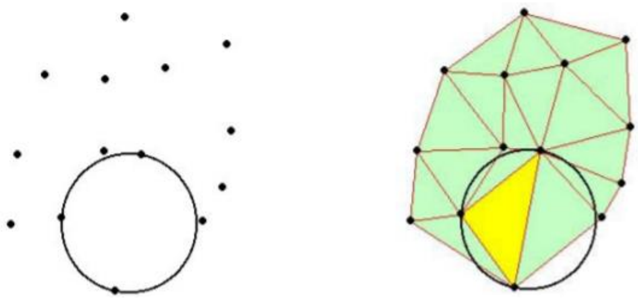


Figure 4: The empty circle criterion that is used for constructing a Delaunay TIN (Verbree & van Oosterom, 2003).

For stream delineation modelling, an adjacent Voronoi diagram is being calculated with the Delaunay triangulation. Then, the triangulation becomes solely a topological connection and the Voronoi cells can be used for the mathematical properties (Gold, 2016). Voronoi diagrams are asymmetrical polygons where any boundary of the polygon is closest to the sample point than any other sample point (Gabrisch, 2011). Figure 5 shows the relationship between the Delaunay triangulation and Voronoi diagrams. The far-right of figure 5 is also called a dual Delaunay triangulation. The Voronoi diagram is used for the generation of the topology and volumetric calculations, while the triangulated surface is the basis of the terrain model (Gold, 2016). This combination is explicitly useful for stream modelling as the edges of the triangles share the same spatial relationship as the adjacent Voronoi polygons created from the same nodes. Both triangulations and Voronoi diagrams are suitable to model terrains elevations. Thereby, this method maintains the quality of the data more sufficiently than alternative interpolations to rasters (Gabrisch, 2011).

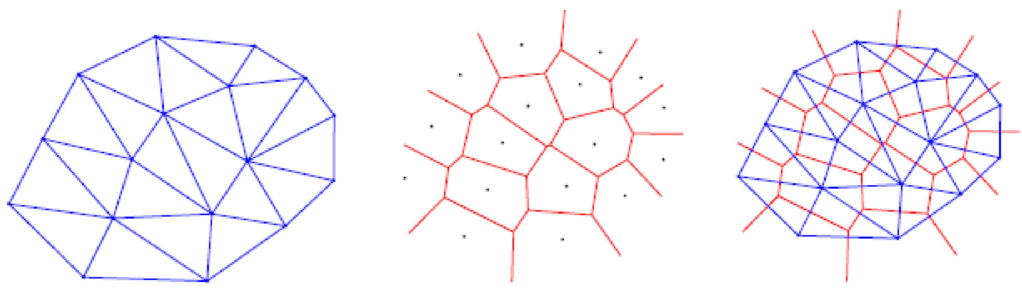


Figure 5: Relation between Delaunay triangulation and Voronoi diagrams (Gold, 2016).

A constrained Delaunay triangulation generalises the regular Delaunay Triangulation by forcing specific segments, usually defined by break lines, into the triangulations. Along the break lines, the Delaunay rules are ignored. In areas with low topographic relief, this method is used to guarantee that important characteristics, such as streams, lakes or other objects in a landscape are maintained. With that, the triangulation consists predominantly of both well-shaped triangles and the topology is accurately mapped; although misshaped triangles could occur at the locations of the break lines as they force the triangulation to conform to these break lines (Marsh et al., 2018). Other 2-dimensional triangulations consist of poorer shaped triangles and are, therefore, less suitable for mapping elevation. Furthermore, contour lines can be generated from point clouds. Contour lines are lines constructed over areas based on equal elevation. By using interpolations, the intermediate elevation values, between the sample points, are calculated to create the contour lines. The contour lines simplify terrain contours and describe geomorphologic characteristics, such as ridges in surfaces accurately (Ai, 2007). By including these contour lines in a Delaunay triangulation as a constraint, these geomorphological characteristics can be mapped more accurately. More specifically, for LiDAR data, contour lines can be used to distinguish ground points from non-ground points in the dataset (Z. Wang et al., 2018). Ground points are considered as all the points of the earth surface and non-ground points are the points that are above this surface, such as objects and trees. Despite the better modelling of small details in the surface, a constrained Delaunay TIN is likely to consist of flat areas along the contour lines (de Azeredo Freitas et al., 2016). The points along the contour lines have the same elevation value and because of this constraint, the triangles can become very narrow with a negligible slope. This is considered a disadvantage when modelling a stream, based on elevational differences. Furthermore, in the case of AHN3, ground and non-ground points are already distinguished (AHN, 2019). Thus, this is a step that can be skipped as the ground points are already classified.

Finally, numerous advantages of TINs can be identified. First, triangulated datasets maintain the source data sufficiently and this results in lower errors in the elevation models (Gabrisch, 2011). Furthermore, the triangulations are very accessible tools and the resulting datasets are very efficient. The combination of the high accuracy compared to the raster DEMs and the efficient data structure makes it potentially very suitable for stream delineation tools.

Data Dependent Triangulations (DDTs)

As state above, DDTs take the height values of the input data points into account during the calculation of the triangles. Triangulations constructed from 3D point clouds are optimal if they consider the z-values of the points during the triangulation (Hjelle & Dæhlen, 2006). DDTs can be implemented in different manners. In fact, any type of triangulation could be used for a data-dependent approach by maximising or minimising the cost-function that expresses the properties of the surface (Verbree & van Oosterom, 2001). The significant difference between the different methods can be illustrated by the following figures (figure 6). Because of the urge to optimise the triangles, the representation of the surface is strongly influenced by the Delaunay representation. This is shown in figure 6: on the left side, a Delaunay triangulation is performed and on the right, a DDT is shown (Dyn et al., 1990). As is shown, in contrast to the right visualisation the triangles are well-shaped on the left side. However, the ridge in the area is more smoothly visualised than with a Delaunay triangulation. Figure 6 also illustrates the most important difference between planar (Delaunay) and data-dependent triangulations: planar triangulations are not satisfactory in areas with a steep slope, for example at areas with steep ridges (Kolingerová, 2004).

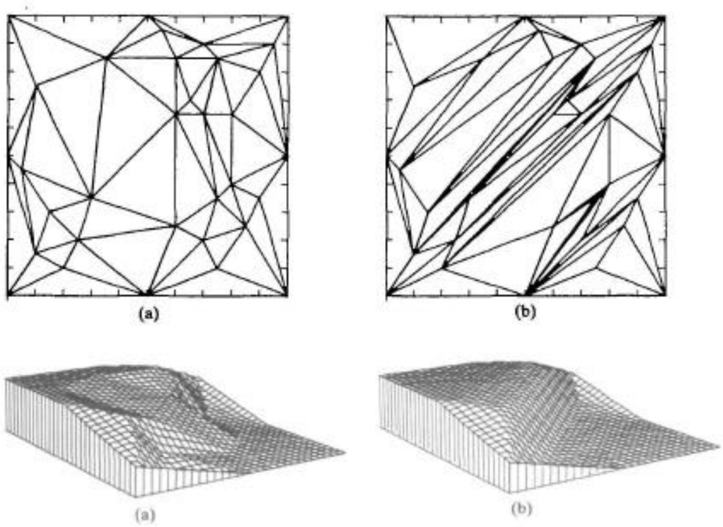


Figure 6: A Delaunay triangulation (a) versus a Data Dependent Triangulation (b) (Dyn et al., 1990).

Rodríguez and Silveira (2017) implement a DDT with higher-order Delaunay triangulations. By using small order triangulations, under three, the accuracy of the terrain models already increased as the root mean square error declined. Creating constrained Delaunay triangulation is also a DDT approach. By adding the contour lines of important topological features to the scope area, these important features can be included in the triangulation (Pfeifer, 2002; Verbree & van Oosterom, 2001). This method relies on an accurate definition of the contour lines. This, however, sometimes is difficult since these lines cannot always be defined as straight lines or the lines are deduced from interpolations of the same input data that is used for the triangulation (Pfeifer, 2002). Accurate constrained Delaunay TINs can be created if the point dataset is from another source than the contour lines dataset. There are many different approaches to triangulate points and mostly it is shown that DDT performs better for terrain models. DDTs are referred to as 2.5-dimensional because it does take the elevation into account during the triangulation process, but the resulting dataset does not have volume. For stream delineation, it could be convenient to have well-shaped triangles to ensure a more natural path of the stream. The elevation data can also be included here by using the z-values of the points after the triangulation has taken place. For instance by looking at the z-values of the points or by assigning a slope to the edges of the triangulated dataset. Lastly, pure DDTs do not perform well on terrain surface data because it results in very sliver triangles and artificial plain areas could be created as local and global criteria could disregard specific geomorphological features of the surface (Verbree & van Oosterom, 2003). The narrow triangles potentially result in plain surfaces in the data and the pure DDT struggles to map ridges and edges.

Tetrahedralizations

To create a 3-dimensional representation of a surface, a spatial decomposition model is defined as all the 3D locations that are closest to a geometric object than to any other (Gold, 2016). There are different approaches, such as the facet-edge, augmented quad-edge or dual half-edge implementation, but the Delaunay tetrahedralization is the most common approach as a result of its simplicity combined with its ability to model volume (Gold, 2016).

A Delaunay tetrahedralization creates a tetrahedral mesh in a 3-dimensional space that discretely represents the space continuously. By finding the closest points in the 3D space, triangles are constructed. This results in a volumetric surface mesh that is created. Therewith, a tetrahedral mesh of the convex hull of its vertex set is created (Si, 2020). An example of these meshes are given in figure 7. The figure shows that a smooth surface reconstruction is created by the algorithm.

Figure 7: a surface mesh constructed with a Delaunay tetrahedralization (Si, 2020).

Tetrahedralizations are capable to interpolate big point cloud datasets to a 3D representation of the surveyed surface. Because of the full 3D approach, this can be considered as the most accurate approach to reconstruct continuous surfaces. Therewith, this method is explicitly suitable for the modelling of overhangs, caves or other complex objects in space (Pfeifer, 2002). Herewith, a smooth surface is guaranteed and the errors are minimised. However, these algorithms, similar to the aforementioned methods, do not take the curvature of surfaces between survey points into account (Si, 2020). Furthermore, these methods are demanding from processing capacities and there are also no ready to use and easy to implement tools available. Lastly, for stream delineation approaches, it is not required to take the volumes into account as the water flows over the upper surface.

2.2.5 No conversions

Aerial LiDAR scanning results in point cloud data. In this chapter, different methods were discussed to interpolate this raw point cloud data to continuous surfaces. However, the raw data also represents the surface in a sampled manner. The characterising high point-density of these point clouds make this discrete representation sometimes sufficient for analysis. Thereby, no conversion of the data means that there is no data loss occurring because of mathematical transformations during the interpolation. The originally measured data is directly being used for the analysis. Anderson and Ames (2011) showed that direct use of raw-point cloud data results in lower errors in the model than when using a gridded raster, retrieved from interpolation of the raw point cloud. By using the z-values of the point clouds, a 3-dimensional representation can be given.

Although the big advantage of assured data preservation, the surface represented discretely. To model the inherent topological relationships between the different points, a continuous surface would be more suitable (Gabrisch, 2011). Additionally, similar to previously discussed data preparation methods, there are no ready to use tools available for analysis with raw point clouds, although the points can be converted to vector points to enable available software to read and analyse the features. Also, the point clouds are demanding from processors and memory because of their huge data density. Although there are filtering methods to compile the data, this could be not desirable, because this also results in data loss. Thereby, for a stream delineation, it is desirable to only use the ground points of the LiDAR retrieved point cloud, but these filtered datasets are still very bulky. Nevertheless, the high data density and the maintenance of the original values of the data could be the fundament for highly accurate analysis.

2.3 Data preprocessing

For raster-based analysis, this step normally occurs after the data preparation phase. Data preprocessing is needed to fill local sinks and pits to assure a continuing flow (Zhou & Chen, 2011). However, point cloud preprocessing sometimes is needed before data preparation to optimise or filter the LiDAR retrieved point cloud to improve the processing capacity and accuracy of the model (Hui et al., 2016). In the case of AHN3, some steps are undertaken to improve the accuracy of the data. For example, GPS-corrections and different flight strips are compared to verify if the data matches. Furthermore, filtering processes are being executed to classify the data to recognise the type of measured object, for example, buildings or vegetation. Lastly, quality checks on point-density, the division of the points, 3D-locations and checks on the filtering are performed (AHN, 2020). These intensive preprocessing methods assure a high quality of the raw point cloud data. Therefore, the quality of this dataset can be considered as sufficient and normally, no further preprocessing measures are required.

To conduct a raster-based stream delineation, a DEM is calculate the flow directions. However, local depressions may impede the continuity of the stream, because the flow directions all point at a specific pixel, or group of pixels, in a depression. These sinks are pixels with the same, or a lower elevation than surrounding cells (Aziz et al., 2020). Sinks appear in valleys, divergent topographical areas and they can also occur because of errors during the creation of the raster. Therefore, to ensure continuity of the stream, sink filling algorithms are normally applied before conducting raster stream delineation analysis. These algorithms generally fill the sinks by using the elevation values of proximate pixels to smoothen the surface of the DEM (Aziz et al., 2020). Commonly used methods use a minimum slope as the parameter to determine what steepness on depressions are filled. These parameters can be adjusted to optimise the results of the sink filling. Mostly the sink filling algorithm of Wang & Liu (2006) is an accessible and accurate tool to fill sinks. Their tool is also available in QGIS. Lastly, it is important to check the DEM whether or not there are no data pixels before filling the sinks. No data pixels are voids in the data and cause non-existent elevation if a sink filling algorithm is used on it. Therefore, the voids must be filled prior to further steps in the analysis (Aziz et al., 2020). However, in a LiDAR retrieved DEM, it is not likely that voids are present because of the great data density of the original point cloud. Thereby, if the interpolation was executed properly, all the gaps in the surface area should be filled with a value.

2.4 Processing

In the case of stream delineation processes, the processing phase is the part where the actual stream delineation algorithm is being executed. Based on the choices made in the previously described phases. The type of analysis is contingent upon the input data that is being used for the research. Therefore, the results of the research are dependent on the type of data that is being used. As described, in the case of stream delineation mainly three input data types can be distinguished: raster, tessellations and raw point cloud data. This chapter reviews different dominant methods to determine flow direction by category.

2.4.1 Raster based methods

Besides the previously mentioned pitfalls of raster-based stream delineation in areas with low relief, such as the loss of information during the interpolation, limitations by pixel size and the restriction of their eight-cardinal directions, this method also offers opportunities (Gabrisch, 2011).

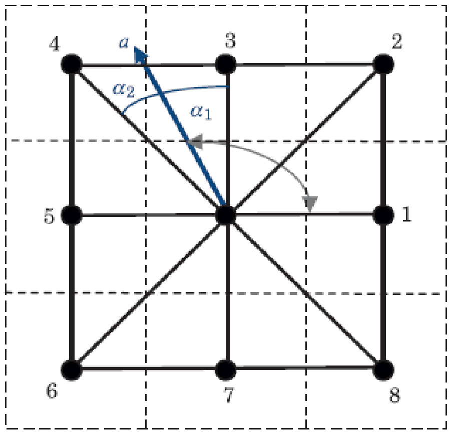


Figure 8: Conceptual approach of the D∞ algorithm (Szczepaniak-Kołtun, 2015).

Raster based approaches are very well-known and many pre-made software packages with good documentation are available. Processing methods to determine flow direction coming with advantages and disadvantages are listed by Gortzak et al. (2020). Their overview is shown in table 1. Predominantly, two methods are distinguished: eight directional (D8) and infinite directions (D∞) algorithms. A distinguishment can be made between approaches that flow to only a single direction (deterministic) or in multiple directions (multi-flow). The D8 can only stream to its eight adjacent grid-cells while the D∞ also takes further grid-cells into account. Based on the values in these cells, the angle of the direction of the stream is adjusted. This is conceptually shown in figure 8 (Szczepaniak-Kołtun, 2015). The stream direction is based on the elevational values assigned to the neighbouring cells. This results in a stream network with many different streams of diverse volumes. Different methods exist to extract the major streams from the smaller streams in this network. Usually, this is done with the Strahler order, where the streams are ordered based on the size. The size of a stream increases as two streams join each other in the stream network; the more streams have joined, the higher the streams are ranked by the Strahler order. The higher the order, the bigger the streams (van der Kwast & Menke, 2019). Another possibility to define the significance of a stream is by using a flow accumulation algorithm. Therewith, the accumulated value of each cell is added up to the cell where it flows to.

Table 1

Raster based methods for determining flow direction

Method	*Author*	*Summary*	*Advantages*	*Disadvantages*
D8 (Deterministic eight algorithm)	O'Callaghan & Mark (1984)	From the source cell, a water particle can flow in a single stream to 1 of the 8 neighbouring cells.	- Watershed boundary and river network are derived easily - Simple and convenient - Widely applicable in hydrology research - Functions well in low elevation areas	- No accurate flow-representation due to only 8 directions (zig-zag pattern visible) - No continuity of water flow in areas with pits
MD8 (Multi-flow directional eight algorithm)	Quinn et al. (1991)	From the source cell, a water particle can flow in multiple streams to all the 8 neighbouring cells	- Truer to life results then D8 due to multiple flows - Avoids concentration of all streams on a single flowline	- Lot of dispersion, even on convergent slopes - Does not function well in low-elevation areas due to multiple streams
D∞ (Deterministic infinity algorithm)	Tarboton (1997)	From the source cell, a water particle can flow in a single stream in infinite directions	- Follows natural stream better compared to D8 and MD8 (less zigzag)	- Does not function in flat areas due to the infinite number of directions
MD∞ (Multi flow directional infinity algorithm)	Seibert and McGlynn (2007)	From the source cell, a water particle can flow multiple streams in infinite directions	- Ideal for topographic index applications	- Mostly useful for overland flow analysis - Not useful in low elevation areas due to multiple streams and infinite directions

Gortzak et al., 2020

For areas with a distinct topography, these methods are particularly suitable. In these areas, the flow-direction is straight forward and there is a clear downward slope in the area. The D8 algorithms are very useful to conduct a stream delineation effectively and easily. However, because of the limited directions in these types of algorithms, the directions cannot be derived effectively in areas with more complex topographical features. The D∞ methods are better suited for areas where finer differences are of importance during the stream direction calculation. These techniques are considered as more refined, but they are still not specifically useful in areas with low topographical relief. The streams that are created will be very dispersed and random if there is no fluent and clear topography in the area (Yang et al., 2017). Furthermore, by using raster-based approaches, a good path can be found. However, van Bemmelen et al. (1993) argue that the exact solution can never be found with raster-based approaches and that it will always be an approximation. For example, a 4-connected raster stream is approximately 41 per cent longer than the actual, optimal length (van Bemmelen et al., 1993). The stream model is always dependent on a specified number of directions to stream to. As stated before, a limitation of several directions result in a deviating and cascading stream, which is not an accurate representation of the natural water flow.

2.4.2 Tessellation based methods

As stated before, flow direction modelling can also be performed using tessellated datasets. Creating triangulated data is a convenient method to create tessellated networks. Usually, the original input dataset is a point cloud dataset. These points are placed in a 3D space and are the vertices of the triangles. These vertices are connected with the edges. Then triangle-shaped polygons are formed that are called the faces of the polygons. A combination of these different facets of a triangulated dataset represent the surface area continuously. Normally, in triangulated representations of surfaces, an elevation attribute is assigned to the faces based on an interpolation of the elevation values assigned to the vertices. These interpolated values can be used for stream delineation modelling. However, this results in an analysis with interpolated data and will be very similar to the raster-based approaches. Therefore, for flow direction calculations, alternative methods with TINs can be adopted. These methods are discussed.

First, Voronoi diagrams are used for creating a flow direction model. Dakowicz and Gold (2007) created a finite runoff model with Voronoi cells that function as buckets that are filled by water based on stream direction information from a Delaunay TIN. The flow is modelled iteratively by examining the movement of the water from Voronoi cell to cell, based on the direction of the flow, assigned to the TIN edges. The water is accumulated in each Voronoi cell based on the inflow and outflow. The outflow in the higher elevation points will be higher than the inflow and the other way around. They conclude that the model can accurately create a global flow model, but it is indicated that the method is complex and the processing performance of this model is particularly low. Other stream delineation algorithms with tessellations can be seen as a walking algorithm over the triangulated network. For example, the edge lines can be used to identify the stream flowing from point to point and to find the path downward (Gabrisch, 2011). Depending on the characteristics of the algorithm the choices for the route can differ (F. Zhang et al., 2018). Using this network for the analysis results in less data loss because no interpolations take place in the process. Potentially, this makes this approach more precise as data is not generalised, but all the original data is being maintained in the vertices of the network. This chapter discusses different stream delineation methods, based on triangulations. First, some assumptions are considered that are required for successful stream delineation (Qu et al., 2014):

1. The flow follows the direction of the steepest slope;
2. The flow continues until it gets into a depression or crosses the research area boundary;
3. For every node, a unique direction with a steepest slope can be identified;
4. Every triangle has a maximum of one edge where the water flows out.

Gabrisch (2011) uses the vertices and the edges to calculate the flow direction on a Delaunay TIN. Using the line distance between the vertices, a slope can be determined. The directions are the steepest descent path from each vertex to its adjacent neighbouring vertex. Based on the flow direction, the natural stream in the scope area can be delineated. The streamflow lines are generated over the edges from vertex to vertex and get a slope attribute value assigned. The flows can globally be hierarchically categorised based on its steepest descent. In this algorithm, the flow direction is directly based on the slopes between the original LiDAR point cloud data. Therefore, this method can be considered accurate and reliable because it maintains the integrity of the raw point cloud data. This method examines the streams of the flow globally in an area with low relief and is presumably more accurate than alternative methods as the TIN surface model preserves the original points of the point cloud. However, a risk is that the forced flow over the edges lead to a rugged and unnatural flow path. Another method where the stream is delineated over the edges of the triangles is suggested by Qu et al. (2014). They propose two methods, where they present the flow direction based on the triangles gradient as the most suitable for surface water runoff modelling. Alternatively, they present a more classical method where the gradient of the edge decides the flow direction. They argue that this type of analysis is mostly suitable for channel runoff modelling. However, this method struggles to effectively handle minor local elevational differences, which results in possibly random local flow direction. Besides, this method is a very local approach, where not the whole surface is taken into account. Specifically, in areas with low topographical relief, the water may flow in multiple directions. This could result in an insufficient description of the actual flow of the streams in an area (Qu et al., 2014).

Besides the above-described methods, the flow path does not necessarily stream from vertex to vertex over the edges. It is also possible to direct the flow through the triangles from edge to edge or from edge to vertex (F. Zhang et al., 2018). This alternative routing can be useful when triangles are not perfectly shaped to maintain the natural gradient of the stream. Thereby, if the algorithm is not constrained to only streaming over the edges, the algorithm is less restricted for the direction it flows to. Furthermore, it is possible to direct the stream over the faces instead of over the edges. Then, the flow direction is determined by the gradient of the triangle. The gradient of the triangle can be accurately established by using the elevation values of the three corresponding vertices (Rheinwalt et al., 2019). This method lets the water drain into each other via the flow lines. This drainage model also delineates water streams, based on the flow direction.

De Azeredo Freitas et al. (2016) propose a method to delineate streams in flat areas by using surveyed contour lines as input data. A constrained Delaunay triangulated elevation model is created with the contour lines, extracted from sample-points. They argue that the constrained Delaunay assures the maintenance of important characteristics and features of the terrain surface. However, the contour lines cause completely flat areas in the constrained TIN. Self-evidently, this must be avoided if the model for runoff modelling is based on elevational differences. Therefore, their method removes completely flat areas by adding artificial points to the data, to ensure continuity of the stream downward. The flat areas are removed by inserting new points on the so-called critical edges of the triangles. The critical edges are the edges that occur in two cases: (1) edges connecting non-consecutive points on the same contour line or (2) edges connecting points on different contour lines of the same height (de Azeredo Freitas et al., 2016). The height of the new points on the critical edges are linearly interpolated between the elevations of the initial and final points. The initial points are the starting points, defined by the triangles with only one critical edge. This method traces the stream flow path of steepest descent with the flow direction defined by the gradient vectors of the triangles. This analysis is particularly useful when there is a limited availability of input data, specifically contour lines, for the stream delineation analysis. Because this research works with high-quality data, this method would not be a logical choice as a lot of available data would not be utilised.

De Azeredo Freitas et al. (2016) note that this analysis, both with raster as with TIN are very local operations. Specifically, in areas with low topographical relief, this could cause problems, because the stream delineation could get stuck in a local topographical depression. As mentioned before, this phenomenon is called a sink and this is a location where there is no lower neighbour for the specific location. These sinks can be handled during preprocessing, but can also be considered during the execution of the algorithm. Traditionally, the sinks are filled using a smoothening algorithm that locally increases the elevation value of the sink. De Azeredo Freitas et al. (2016) suggest to remove these pits with an interpolation of the Z-value of the points in the TIN. Rheinwalt et al. (2019) suggests a method to handle the sink problem in landscapes by carving the sinks. Therewith, the water is directed away from the bottom of the sink, instead of filling the sink. If the algorithm comes into a sink, it causes circling in cycles. If this is being identified, an artificial tunnel can be created to the lower values behind the higher triangles around a specific point. Therewith, the locality of the algorithm is being broadened by not only looking only at the adjacent vertices, but also to further points until a path downward has been identified. The approach of Rheinwalt et al. (2019) is conceptually shown in figure 9. The turquoise arrows represent the stream cycling in a sink, the pink arrows show the tunnels that are created when running into the sink and the blue arrows represent the flow direction (facet-flow network).

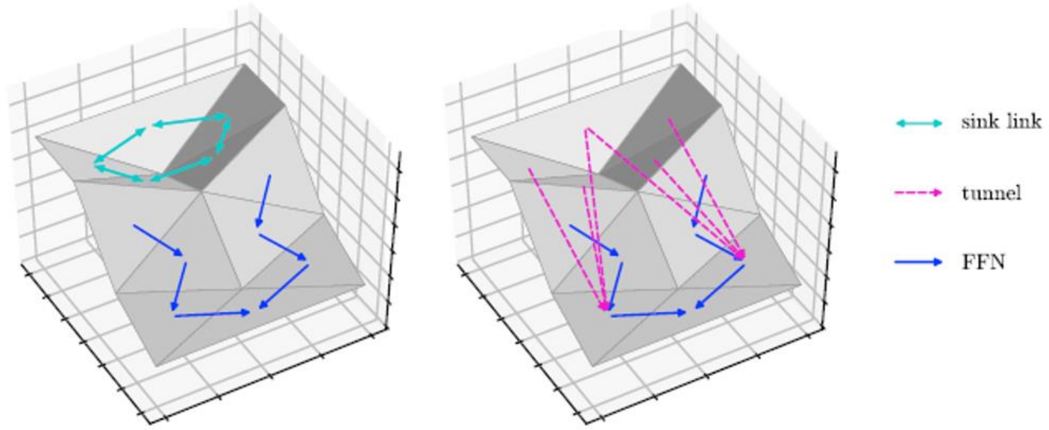


Figure 9: conceptual visualization of handling sinks by increasing the locality of the analysis (Rheinwalt et al., 2019).

Concludingly, it can be stated that applying a stream delineation algorithm correctly is more complex than the conventional raster-based approach, but it most likely results in more accurate results. By applying a stream delineation algorithm that walks over a network of edges, there is a risk that the algorithm is longer than in reality because it has to follow the connections between the points. Therefore, because of this long path that is chosen, the sinuosity (the length of the stream to the valley length) could be higher than the actual sinuosity (Anderson & Ames, 2011). A network-based flow algorithm, similar to the method of Gabrisch (2011), seems to be the best fit for this research concerning the requirements of accessibility and accuracy. However, the simplicity of this method could become a challenge. A method where the water flows over the edges of the triangulated network presumably is more accurate than conventional methods and meets the requirement of simplicity. Nevertheless, this method could result in very long streams as it is obliged to follow the edges of the triangles. By allowing it to flow over the surface, through the faces of the triangles, the algorithm tends to get very complex because it is difficult to determine when it has to choose a path over the edge or when it has to flow over the network. The original point cloud data can be preserved and the details in the landscape are better considered in this model. Thus, these methods presumably are more suited for analysis in areas where the topography is more complex, where details in the landscape play a more important role in retrieving highly accurate drainage networks (de Azeredo Freitas et al., 2016; Gabrisch, 2011; Qu et al., 2014; Rheinwalt et al., 2019).

2.4.3 Direct point cloud-based methods

Lastly, methods directly applied on point clouds, without conversion of the data are discussed. Direct methods are defined as methods where the raw point cloud is used as input for the stream delineation algorithm. This section discusses different methods that are identified as direct point cloud analyses. Using direct point cloud data is relatively innovative and therefore, new methods, based on knowledge about the concept of stream delineation, are suggested.

Anderson and Ames (2011) created a stream delineation algorithm, the mDn method, that automatically creates a neighbourhood, divided into several sectors, around an arbitrarily selected starting point. Different parameters can be adjusted: the size of the neighbourhood, the number of sectors and the maximum allowable course change. This method is conceptually shown in figure 10. The average elevation of a sector is calculated based on the LiDAR points that fall within a specific sector. Then, the stream flows to the sector with the lowest elevation and the process is iterated until the algorithm is terminated. This is a convenient method to perform a stream delineation directly with raw point cloud data. The results show a lower root mean square error of this method than conventional, raster-based, methods. Nevertheless, some shortcomings of this approach can be identified. First, during the calculations, the point values are generalised in a mean elevation value; the optimal usage of the high information density, therefore, is not utilized. Moreover, the algorithm is created for ArcView, a GIS-platform that does not exist anymore. This limits the usability and accessibility of this approach. Finally, this algorithm is very sensitive to voids and sinks in the input data. Very careful parameter adjustment is required for this method, otherwise the calculations will terminate. These forced choices potentially influence the results of the stream delineation strongly (Anderson & Ames, 2011).

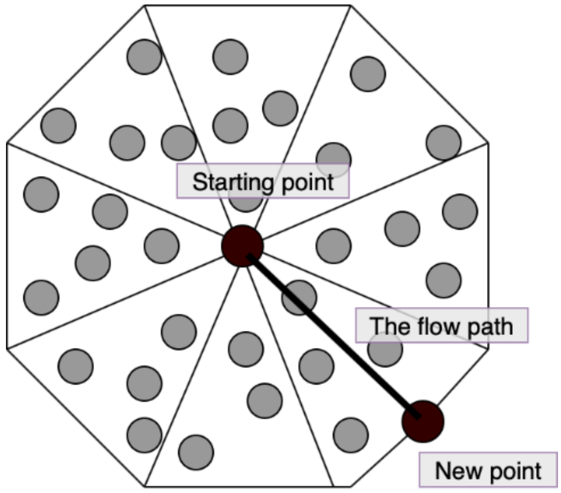


Figure 10: Visualisation of the mDn method. The water flows towards the sector with the lowest elevational average (Gortzak et al., 2020).

To overcome the problem of generalising the raw point-cloud data, another neighbourhood-based approach could be adopted where no averages are calculated. Similar to the mDn method some adjustable parameters are used, but this approach does not use sectors with generalised elevation values. Within the created neighbourhood, the point with the lowest absolute Z-value is selected and the stream is directed to this point. This is based on spatial autocorrelation, where it is assumed that nearer points are more related to each other than points that are further apart. This is an alternative method that is not executed yet for stream delineation. However, Ujaval Gandhi (2020) suggests a method to select neighbouring points using a search radius in QGIS with decision rules. By using the field calculator, a specified relation to each individual point within a search radius can be identified. Also, the K-nearest neighbour algorithm could be used bere. This way the number of neighbours can be specified and the search radius is adapted automatically, based on the input data (Zhu et al., 2016). In the case of a stream delineation, the point within this search radius with the lowest elevation is identified. This method is similar to the mDn method, but it does not use averages. In this case, the algorithm selects the absolute lowest point that is shown within the radius in figure 10. Also, a search angle within this radius can be included. This can be used to push the algorithm forward and prevent it from flowing back. Based on different decision rules the method selects the most presumable point where the stream flows to. These decision rules can be based on characteristics of natural flow and the input data. This method is very local and, therefore very sensitive to get stuck in local depressions. However, the search radius can be easily extended locally if there is no lower point identified within the originally specified neighbourhood. Furthermore, this method is restricted to a single stream delineation as it is not possible to identify multiple paths at once. The locality of the algorithm results in a stream delineation from point-to-point and the multiple streams over the whole surface cannot be identified herewith. Finally, this method would require high processing capacities because of the high point-density. This could be mitigated by creating spatial indexes of the points before processing them. Also, as this is a local approach, the calculations can be performed locally. This limits the required points to be processed. Therewith, the speed of this algorithm is boosted.

Furthermore, it is possible to calculate the flow, based on the normal vector of the points. The normal of a surface is a vector perpendicular to that specific point and says something about the orientation of these points. The normal can be calculated with different local surface models such as a plain area, a 2D triangulation or a quadric model (CloudCompare, 2020). The method with the plane neighbourhood is robust to noise but edges and sharp corners are not taken into account accurately. The quadric method is very suitable for curvy surfaces and the 2D triangulation is weak to noise but is good with sharp edges. As represented areas typically are not curvy and because the noise in the dataset is relatively small because of the high quality, the 2D triangulation method is presumably most suitable for this research (CloudCompare, 2020). With the 2D triangulation method, the normal of a point or triangle is determined based on a Delaunay triangulation.

The normal vectors can be used to identify the direction of the flow at a specific location based on the Z-value. This research focusses on the Z-value orientation of the vector. The flow streams over the surface where the Z-dip is most significant. The orientation of these points can be used to identify the stream direction of the water. This direction can be used for individual points (from point to point) but also the direction of the flow over a triangulated network can be defined with normals, instead of the slope as the determinant of the direction. The stream can be delineated from an arbitrary starting point and flows towards the next point where it aims to, for example over the edges over a TIN. This is a local application as the stream flows from point to point and streams for the whole surface are not calculated. If looking for a more global application, this approach can be useful when combining it with Voronoi diagrams. An approach, similar to the previously described bucket method (Dakowicz & Gold, 2007), is very applicable with the normals at each point as the water is poured into the Voronoi cells and filled like a bucket based on water height and direction at each point. Dakowicz and Gold (2007) determine the direction of the stream with a Delauney TIN and the assigned slopes, but the directions at each point could also be determined with the assigned normals. This method would result in a global flow direction model and this approach would not meet the requirement of simplicity of the model as this method requires high-level programming and performance issues would occur here. In these methods the original points are used and the normal vectors can be calculated in freely available software: CloudCompare. The normal vector approach and the triangular network approach are similar because the orientation of the points, based on the normals, will be approximately similar to the direction to the edge with the lowest gradient. Also individual points can be used for the direction. The normal vectors of the points provide useful information about the orientation and direction of the surface at a specific point. Because the normals of the points can be easily computed, it is a promising approach. However, to use the normals, a TIN or other surface representation is needed to determine the path. Thus, this method is not considered as an entirely direct approach as the transformation of the point cloud data is needed.

Direct point cloud-based methods do have great potential, because the process of stream delineation is simplistic; only processing steps are required since the raw point cloud serves as directly as the input. Hence, the results of these methods presumably will be accurate. However, the currently known approaches harm the integrity of the raw point cloud data, are still not very accessible because of their limited availability or do not take important surface features into account (Gabrisch, 2011). Therewith, the approaches can be very local and at risk to end up in sinks. For instance, this risk should be taken into account with the absolute point neighbourhood-based approach as it is based on very local operations. Nonetheless, the method is relatively accessible, easy to apply and has the potential to be very accurate as it maintains the integrity of all the points.

2.5 Conclusion

This chapter overviews different stream delineation methods. The main focus is to find an accurate, simple and accessible application that maintains the integrity of the raw point cloud data as much as possible. An overview of the strengths and weaknesses of different approaches was given. Based on these insights, three candidate methods were identified. The raster-based approaches do not maintain the high information density of the original data because the data is being generalised during the necessary interpolations. Furthermore, other methods, with an interpolated TIN or with a local neighbourhood that calculates the average per sector, is generalising the points too much to maintain the integrity of the data. Triangulations maintain the data integrity and are particularly suitable because of their efficient data structure. However, the analysis can become complex and not accessible. Vector-based methods are suitable because the stream delineation model is based on a very accurate orientation of the points. However, vector-based methods are also considered as complex and not accessible, yet. Lastly, methods, based on the absolute values of points are suitable because of the utilisation of the original point cloud data. However, problems can occur when looking for an algorithm with sufficient computing speed. The three candidate methods are listed:

1. TIN network-based;
2. Normal vector-based;
3. Absolute point-based.

The next chapter discusses these three candidate approaches more elaborately. This theoretical review functions as the fundament of the methodology for this research.

3 Methods

This chapter describes the methods that are adopted to delineate streams in a selected scope area. Since the scope area has relatively low topographical relief, this research aims to adopt a more accurate algorithm, as the details in the surface matter and data loss from the raw point cloud is not desired. By creating four prototypes, the most suitable method for this research was selected. First, this chapter presents the methods for two Delaunay TIN-based approaches, a normal vector-based approach and a direct point-based approach. Based on the results of these prototypes, the presumable best method was selected and further developed towards a well-functioning algorithm. This process is visualised in figure 11 and gives an answer to the second sub-question: *"What (combination of) approaches with direct point cloud analysis is most suitable to delineate streams in an area with low topographical relief?".*

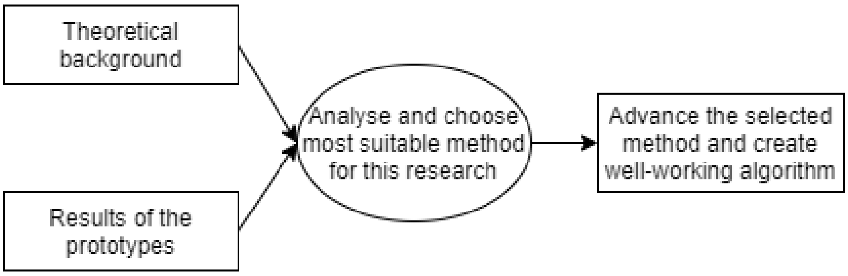


Figure 11: Methodological approach for selecting the most suitable approach

3.1 Prototypes

This paragraph discusses the methods to create four prototypes that possibly meet the main criteria of this research to create an accessible, simple and accurate stream delineation algorithm in areas with low topographical relief. The TIN-network approach is a process where the water flows over the edges of the TIN, based on the slope assigned to each edge. Alternatively, the water can be directed based on the normal vector of the vertices, attached to edges. The normal vector-based approach uses the directions of the individual points to determine the flow direction per point in the research area. The absolute point approach looks at the absolute points within a search radius of the reference point.

The four methods are executed in a very small test area with approximately 4000 LiDAR points. The used research are both consist of more than 100 times more points (see table 3, page 53The prototype tests are executed in QGIS, using existing libraries and basic programming in PyQGIS. These prototypes are used to get a better understanding of the performance and specifications of the different approaches to select the best method for this research. Because the prototypes are executable in the open software of QGIS, it is assumed that all the methods meet the requirements of accessibility. The simplicity and accuracy has been examined based on the preliminary results of the prototype algorithms. To examine the accuracy of the stream, a reference stream, provided by the local water authorities is used. The selected area is a part of one of the scope areas, that are lightened out in chapter 3.4. This area is selected, because it is representative of the whole scope area, contains a curving stream and also contains areas with relatively low data density. This low data density is a result of the filtering of ground points that are used for the stream delineation. The different approaches that are identified are essentially the same, because all the streams identify the fastest path down, based on the elevation and their relation to neighbouring points. The difference between the methods is how this relationship between the different points is defined. The distance that the water can flow to neighbouring points and the number of neighbouring differs per approach. Also, the determinator for the flow direction differs, where the first approach is based on the slope that is assigned to the edges of TINs, the second on is based on the normal vectors, based on triangulation. Thirdly, the absolute elevation values in a specified neighbourhood are used.

3.1.1 Constrained and normal Delaunay TIN network-based approach

As stated before, an approach that maintains the integrity of the data is by creating a triangular network with a Delaunay triangulation. The created dataset functions as a network of vertices (the original points) and edges (the connection between these points), where a slope is assigned to the edges. The Delaunay TIN is created by using an existing QGIS algorithm. This library provides an algorithm that efficiently creates Delaunay TINs (*QGIS API Documentation*, 2020). This tool also exports edges of the triangulated dataset to vectorised line-segments. The edges are now individual features and values can be assigned to them. The edges must consist of a slope gradient as the water flows to the point over the steepest slope downward. The slope of the edges is calculated in a three-dimensional surface with x, y and z coordinates assigned to each point. The slope is calculated with the following formula:

$$Slope = \frac{Rise}{Run}$$

The rise is the difference of the elevation (z) value of the points, while the run can be calculated using the Pythagorean theorem:

$$Run = \sqrt{(\Delta y)^2 + (\Delta x)^2}$$

Once the slope is calculated, it can be assigned to the vertices. These values can be used for the flow direction calculations, while the steepness of the slope determines the direction and the significance of the flow. Now, the local flow direction for each edge for the whole surface is defined. For this prototype, these stream directions are used from point to point. An arbitrary starting point is selected upstream as the beginning point for this algorithm. The prototype algorithm is executed manually. To make sure the stream goes forward, a maximum course change was integrated by restricting a course change angle between -88 to 88 degrees.

This method is also executed on a constrained Delaunay TIN. This dataset is created by using the contour lines, extracted with a normal triangulation of the points, as a constraint during the triangulation. After that, the above-described method is adopted on this dataset. For the creation of the contour lines, a contour interval of 0.4 meters is used. Linear interpolation is used to treat each triangle as a plane. The contour lines change from direction when they enter the adjacent triangle (Esri, 2021).

3.1.2 Normal vector-based approach

Furthermore, the normal vector-based approach is suggested. The normals of the points are extracted using CloudCompare, open-source software that is designed to work with point cloud data. The normal vectors are based on triangulation and indicate the Z-orientation of the points. During the creation of the prototype, multiple approaches were explored based on the direction of the normals. The extraction and the orientation of the normals is based on a triangulated approach. Many challenges were faced and it was not possible to create a prototype that met the requirements of this research: accessibility, simplicity and accuracy. However, in the results section, the normal vectors were visualised and reviewed to provide an understanding of the potential of this method.

3.1.3 Absolute point-method

Lastly, a method that uses the absolute points, without data conversion, was lighted out. In the case of this prototype, the neighbourhood was set to 3. With the field calculator, the lowest neighbour of a specific point can be identified. Using a script in PyQGIS, a line is drawn between an arbitrarily selected starting point and the lowest neighbouring point. This process was iterated until the lowest point in the area was reached. For this prototype, the lowest neighbour of each point in the dataset was identified.

3.2 Absolute point-method

An advanced neighbourhood-based approach looking at the absolute values of points was created for this research. The stream is modelled from an arbitrarily selected starting point towards the lowest neighbour in a specified neighbourhood. This process is iterated until the model reaches the lowest point in the research area, the outlet. The accuracy of this method was validated by comparing it to the results of a conventional analysis using a D8-algorithm and by logical interpretation. All the analysis steps were conducted in QGIS. This open-source software is particularly suitable because everybody has access to it and works well with point cloud data. Besides, there is a wide availability of ready to use tools, that can be integrated into this method without difficulties. These tools are stable and tested and therefore, the expected number of errors that occur within these tools is negligible. It is not only possible to view the point clouds, but it also is possible to easily convert them to other data types, for instance, Delaunay TINs or vector point datasets. If the available tools need adjustment or no tool is available, the Python extension of QGIS can be used. From here, this chapter describes the different steps that are taken during this research. First, the analysis design is presented. Then, a suitable scope area is identified. Therewith, an answer to the third sub-question was be provided. Thereafter, the data preparation, preprocessing and processing methods for the newly suggested approach are described. Lastly, the validation methods are described more precisely.

3.3 Analysis design

In previous sections, it is explained that the process of stream delineation requires different steps. This research focusses on the data preparation and the processing of the data. The newly designed method should be an accessible, simple, accurate and well-performing tool. Based on the literature review, a choice was made for a neighbourhood-based approach, where absolute points in an adjustable neighbourhood determine the flow of the water. This method will be referred to as the absolute point-method. AHN, LiDAR retrieved, point cloud data was used as input for this research. The ground points are already filtered and classified (AHN, 2019). The main advantage of this approach is that it is possible using existing libraries in QGIS and that the original point cloud data is directly used. The ground points are used as the input for the absolute point-method. After selecting an arbitrary starting point, the lowest neighbour of this point is identified. Thereafter, a line is drawn between the selected starting and ending point. This process is iterated until the outlet of the area is reached. The algorithm is more elaborately discussed in paragraph 3.7. The designed model was validated by comparing it to a conventional, raster-based, stream delineation and an existing reference stream. Also, the vertical performance was examined by looking at the longitudinal profile. The validation model is executed with the D8 algorithm. An IDW-interpolated DEM is used, as AHN provides an advanced 0,5m raster DEM created using a squared IDW interpolation (AHN, 2019). A discrepancy between the results of the different approaches was expected, as the input data and processing methods differ drastically. Theoretically, the D8-method performs less in areas with low topographical differences. However, this method is very accessible and performs better in low elevation areas than alternative conventional approaches (Szczepaniak-Kołtun, 2015). Logical interpretation, based on the elevation models and their sinuosity and Root Mean Square Errors (RMSE) were used for validation. The RMSE is calculated by the root mean differences between a selected

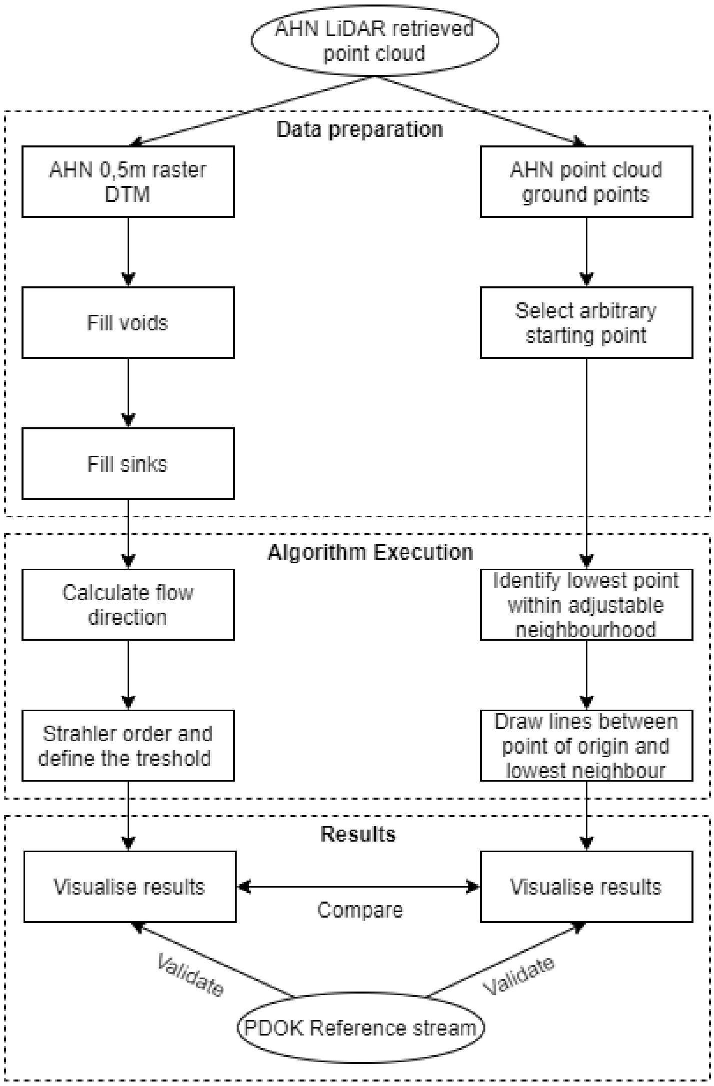


Figure 12: Workflow of the stream delineation processes.

reference dataset, provided by PDOK and considered as most accurate, and the delineated streams by the different algorithms.

This chapter elaborates on how an algorithm was created to accurately delineate streams in an area with low topographical relief. The analysis can be subdivided into three components. First, the data preparation, then the algorithm execution and finally the model is validated when the results of the calculations are available. Figure 12 shows a flowchart of this process. The design of the stream delineation algorithm that is developed for this research is shown in figure 17 in section 3.5. First, the research area and the source data is discussed.

3.4 Research area

The main goal of this research is to provide a proof of concept of a well-working algorithm to delineate streams in an area with low topographical relief. Therefore, The research area must meet some requirements:

- The scope area is a predefined (sub)-catchment with a clear outlet;
- The area should be of low topographical relief, yet a pattern from high to low elevation must be present;
- AHN3 data must be available;
- No human interventions to ensure natural flow.

Two research areas are selected. One area with a clearly identifiable topographical elevation difference and a small area with very small elevational differences. In a small research area, the test dataset is smaller and tests can be executed more easily because faster processing is possible. First, the performance of the created algorithm was tested in an area with some relief. Although there is relief in this area, this area still can be considered as an area with low topographical relief compared to other hydrological studies. Low topographical relief can be defined as areas where subtle differences in elevation values appear (Gabrisch, 2011). For this stream delineation analysis, the whole of the Netherlands was considered as an area with low topographical relief; also the hilly areas in Limburg. This research distinguishes two research areas. One of the areas has a distinct topography, while the latter has almost no topographical relief. The selected research areas are small sub-catchments: the Beversbergbeek, shown in figure 13 in the southern part of Limburg and a stream around Olst, shown in figure 15. As is shown in the longitudinal profile graph in figure 14, the surface at the top is 15 meters higher than at the bottom of the stream. In the area with very low topographical relief, the total elevational difference is less than 0.2 meters over a length of 600 meters. Lastly, there are no human infrastructures in the Epen research area that possibly influence the natural water stream. The second research area in Olst, shown in figure 15, has an almost negligible relief. It was expected that the conventional D8-method does not work in this area when examining the longitudinal profile in figure 16. This very spikey pattern is presumably caused by the very low relief and because the stream is channelised. Also, the big spike, just before the horizontal distance of 400 meters is caused by a culvert that is located there (PDOK, 2020). However, although it is almost negligible, there is a downhill pattern recognizable. If the direct point-method works in this area, it presumably is very effective in all areas with low topographical relief. The two research areas will be referred to as respectively Epen (Limburg, higher elevation area) and Olst (lower elevation area).

A reference stream provided by PDOK was used for validation. This dataset is compiled from datasets from different Dutch water authorities. The most current information is provided via the national geo-register PDOK. It is noted that not all water authorities provide all the data, but this dataset provides the best possible information about the local stream as a reference (PDOK, 2021). This research will refer to this dataset as 'the reference stream' from here.

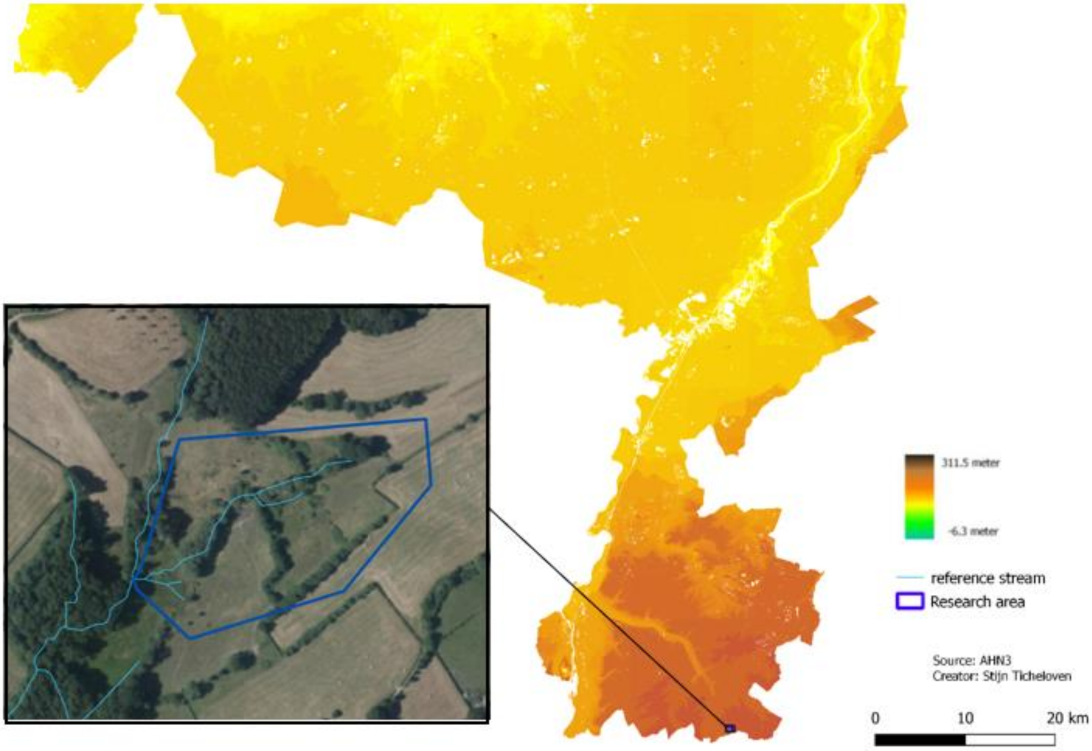


Figure 13: Epen research area: a small water stream in the hilly southern area of Limburg

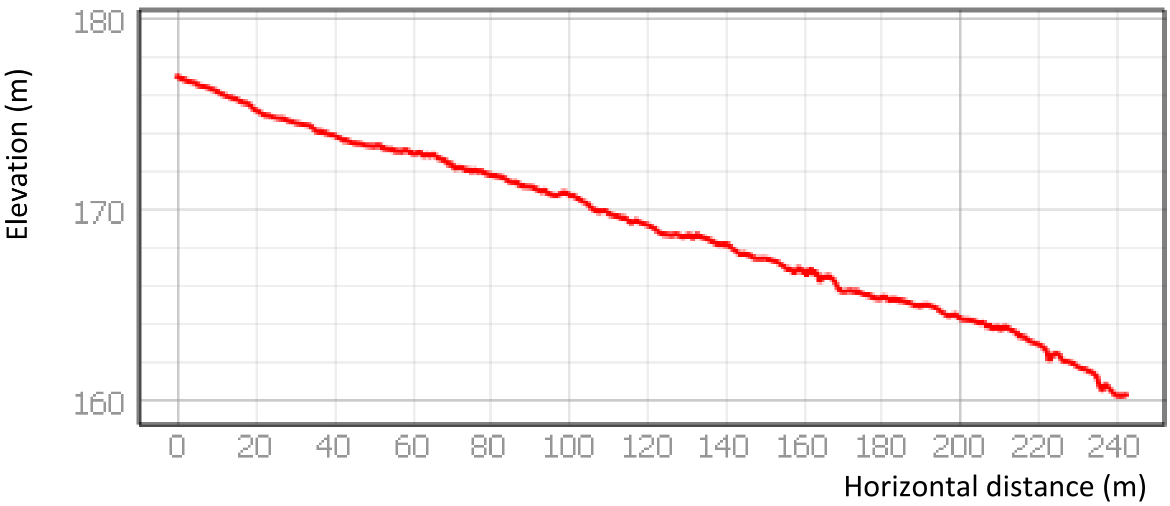


Figure 14: Longitudinal profile of Epen research area.

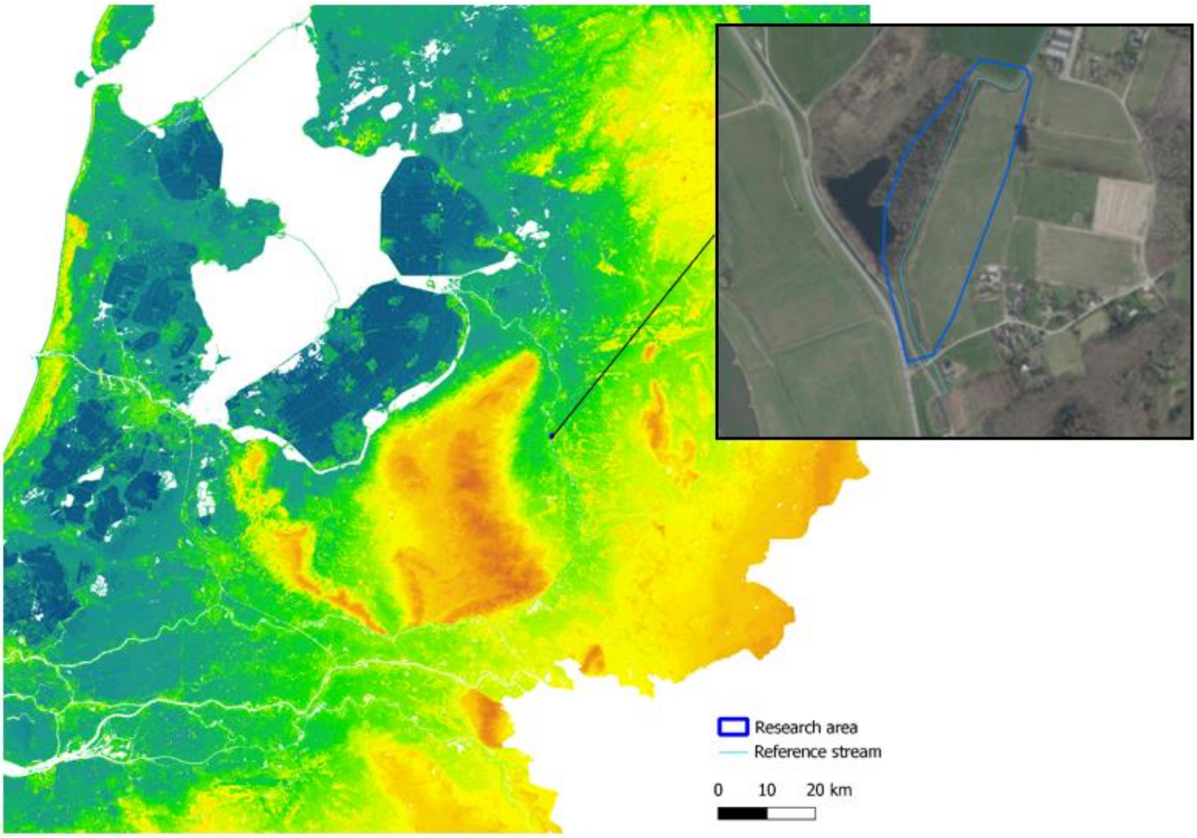


Figure 15: Olst research area: a small water stream in the middle of the Netherlands near Olst.

Figure 16: Longitudinal of Olst research area.

3.5 Source data and preprocessing

This research uses AHN3 LiDAR retrieved point cloud data. As stated in the literature review, raw point clouds from the AHN are already sufficiently preprocessed. As the research area is located in the Netherlands, AHN3 is a very suitable dataset of the Dutch governmental agency AHN that provides high-quality LiDAR data. AHN3 LiDAR data is retrieved with airborne LiDAR. A laser scanner is attached to a plane and transmits laser beams to the earth surface. The beams are scattered back and then analysed by the receivers in the sensor. Based on the calculated travel time of the laser beam, the distance to the earth, or another specific object on the earth, can be determined (Meesuk, 2017). This commonly results in a large dataset that consists of independent LiDAR-points with x, y and z values: a point cloud (Longley et al., 2015).

AHN3 data is gathered between 2014 and 2019 with aeroplanes and helicopters. The point density is averagely between 6 to 14 points per square meter. The point density depends on the type of landscape and the adopted methods for data gathering (AHN, 2019). The Epen scope area of this research consists of 760 095 points, of which 431 404 are classified as ground points. In Olst, this is respectively 988 747 and 686 181 points. For this analysis, the ground points have been used. The size of the Epen research area is 36.536 square meter, while the Olst research area is 110 368 square meter. That comes down to over eleven points per square meter for the Epen research area and over 6 points for Olst. Typically, the low elevation area is less dense with points that the higher elevation area since there is more scattering if there is a less heterogeneous surface. Usually, a DEM with a very high spatial resolution has approximately maximal 0,5m² grid cells (Aguilar et al., 2010). This means that a high-resolution DEM almost consists of two data points per square meter, which is much lower than the corresponding LiDAR elevation model. Furthermore, because AHN3 is calibrated precisely, the height errors are not low. An overview of the errors and the height accuracy is given by AHN and is shown in table 2 (AHN, 2019).

Table 2

Overview of errors and the accuracy of AHN3

Accuracy parameter	*Error*
Systematic error	5cm
Stochastic error	5cm
68,2% of the points have an accuracy of at least	10cm
95,4% of the points have an accuracy of at least	15cm
99,7% of the points have an accuracy of at least	20cm

AHN, 2019

These raw LiDAR retrieved points are classified and only the ground points were included in the analysis. The points, with their corresponding z-values, were not modified any further. Thus, these points are used as input for the final analysis. This presumably improves the accuracy of the calculations, as the originally measured -highly accurate- values are used for the analysis.

3.6 Data preparation

As stated before, this research uses points directly from raw point clouds to minimise data loss. To use existing libraries in QGIS, the LAS-points are converted to multipoint Geopackage datasets. Next to the x and y coordinates, these points consist of a z-value that represents the measured elevation. For this research, the AHN LAS tile is clipped to a smaller area. Since the designed algorithm is a local approach, the research area is clipped to a minimised research area to improve the performance of the model. On the other hand, to apply the D8-method, a square scope area is used. This was done to optimise the result of the performed D8-method. For the absolute point-method, no further data preparation is necessary.

The D8-method inputs an interpolated AHN raster. This raster is created by AHN and is based on a squared IDW of the LiDAR points. The DTM (the digital elevation model with only the ground points) was used. First, the data gaps were filled and thereafter, the sinks were filled by using the commonly known algorithm of Wang and Liu (2006). The filled rasters serve as input to create a flow direction raster, where one of the eight possible flow directions is assigned to each cell. Based on the flow direction raster, a stream network can be identified by using the Strahler order. The absolute point-method is further explained in the next paragraph.

3.7 Processing methods

All the data is arranged to serve as input for the final algorithm that applies the stream delineation. The algorithm finds the fastest way downward and, therewith, simulates how water flows from an arbitrarily selected starting point down. This algorithm uses the absolute elevation values of each point in the dataset. The water flows from a specific point towards the lowest identified point within a, specified by user, search radius. As the lowest points are searched, the streams from upstream downward. If the stream would flow upwards, the stream would be modelled away from the stream because these points are commonly higher than the points in the river bed. The starting point must be chosen by the user of the algorithm. This starting point is arbitrary and it is important to choose this starting point with care. If the starting point is selected sufficiently, the stream should be delineated towards the outlet of the research area. Often, the outlet of the research area is known, and if the stream does not lead towards the outlet, an alternative starting point of the algorithm should be found to optimise the results. This algorithm executes calculations until it reaches the outlet of the area. By default, this is the lowest point in the dataset. However, the endpoint of the algorithm can be changed by the user manually. Finding a suitable starting point for the algorithm is a process of trial and error. Furthermore, this model does not take the sub facet, underground level into account and only considers the water that flows over the top surface. Thus, this model ignores possible infiltration of water during the stream; it is assumed that the water is homogeneously distributed along the surface contours (Rheinwalt et al., 2019). It is supposed that the surface where the water flows over has the same properties and water flow behaves the same everywhere. Besides, the volume of the water was not taken into account in this algorithm. The path towards the lowest point in the neighbourhood determines the direction of the flow in this model.

This model deals with local depressions by extending the search radius for a new point if there is no lower point identified within the specified search radius. Just like the default search radius, the extended search radius is also a parameter that can be adjusted by the user. If also no lower point is identified in the extended search radius, the stream flows to a point with a higher elevation. This is included in this algorithm to ensure the continuation of the flow. A choice for this method of sink handling was made because this method was considered as both accurate and simplistic; two of the main requirements of this research. Other methods that could have been adapted, such as smoothening the data by identifying points with no lower adjacent points, is more complicated and do not meet the initial requirements of this research (Rheinwalt et al., 2019). Moreover, one of the key advantages of this algorithm is the maintenance of the original input data. A smoothening algorithm possibly harms the quality of the original input data as it modifies the elevational values of features.

With the maximum allowed course change, the correct direction of the stream is guaranteed. Furthermore, some topographical rules should be included. For example, the stream cannot cross itself (Anderson & Ames, 2011). Also, a maximum course change of the stream was included to ensure a natural flow of the water. Next to the adjustable search radius, this is a parameter that can be set by the user of the algorithm. Lastly, a projected coordinate system - RD New - was used, to make sure that all the inputs and results are metric. This ensures that all the distances are measured in meters.

As stated before, this method maintains the original points and their assigned z-values. Therefore, this method is useful for stream delineation in areas with low topographical relief. The flowchart of the created absolute point-algorithm is given in figure 17. By that, the stream can flow in an infinite number of directions, dependent on the set maximum course change by the user. Furthermore, this algorithm is very local and only looks at the points within the search radius. The travel distance of a line segment is variable, with a maximum travel distance of the size of the search radius. The closer the lowest point within the neighbourhood, the shorter the line segment that will be drawn. The global flow of water over the surface is not taken into account by this algorithm. Therefore, if a global flow model is desired, this algorithm should be used as an addition to conventional flow models. Specifically, if these conventional methods return errors because of too low elevation differences, this algorithm will presumably be of added value.

This method has three parameters that can be modified by the user: (1) default search radius, (2) maximum allowed course change and (3) extended search radius. The extended search radius applies when the neighbourhood must be extended if no lower points are identified. The starting point has no restrictions. Therefore, the search angle at the first point is always set to 360 degrees. The search angles of further points in the algorithm are based on the angle of the start and endpoint of the previously drawn line. The direction in radians was calculated by using the function atan2(Δx, Δy). Where the delta's are calculated by subtracting the starting point from the endpoint of the previously created line segment. Based on this angle, the maximum allowed course change is added and subtracted to define the search angles for the new point. The values that are returned are between -180° and 180°. As the input must be between 0 and 360, the negative angles were changed to positive angles that fit within the full circle. The maximum allowed course results in that the search radius is being reduced to a half-circle or smaller. Therefore, it actually only really is a 'radius' if the maximum allowed course change is set to 360 degrees.

To improve processing capabilities, the lowest neighbours are only identified for the points where the model streams over. Other points in the dataset remain unused. Moreover, also the lowest neighbour in the extended search radius only is calculated if the lowest neighbour in the default radius has a higher elevation than the starting point of the specific line segment. This is done to minimise the calculations that must be executed, to improve the speed of the algorithm. Correct parameter optimisation ensures accurate results of the stream delineation in areas of low topographical relief. Dependent on the data density, the number of outliers and the characteristics of the surface, the optimal parameter settings can differ. Typically for low relief areas, no drastic course changes of the stream occur, while a more drastic course change can occur in more steep areas. This must be taken into account during the parameter optimisation. Lastly, the algorithm will terminate if it arrives at the lowest point in the research area, which is generally located at the outlet. In appendix A, the PyQGIS code for the prototype is attached.

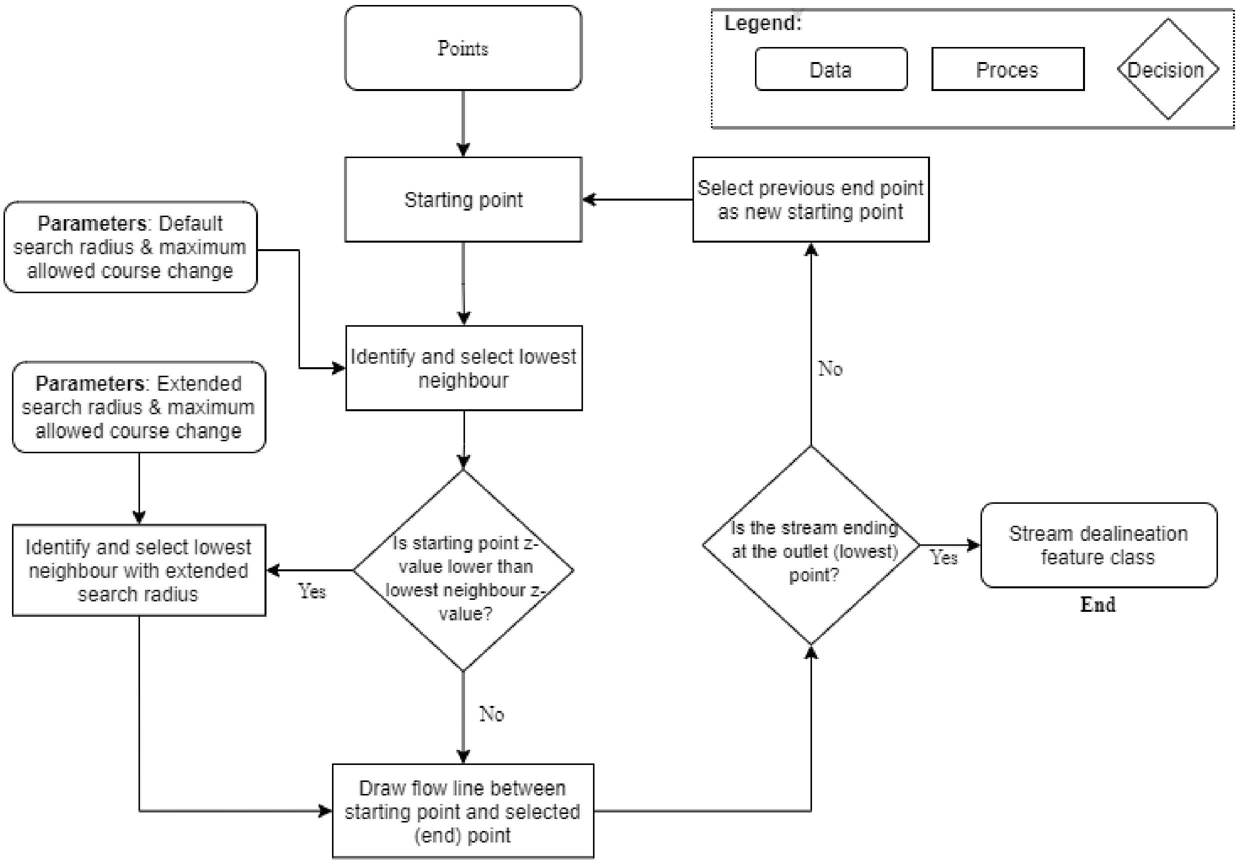


Figure 17: absolute point-method delineation algorithm flowchart.

3.8 Model validation

As explained in the analysis design, the model was validated by comparing the stream algorithm to a conventional D8-method. In the selected research area a reference stream is provided by PDOK, the Dutch governmental spatial data provider. These streams are acquired using aerial photography, terrestrial collection or using large scale topographic maps. This information is continuously being updated (PDOK, 2020). For the model validation, the root mean square error (RMSE) is calculated of the extracted streams, compared to the reference data. The RMSE gives a degree of match between the reference stream and the modelled stream. The better the match, and thus, the lower the RMSE, the more accurate the model is (Anderson & Ames, 2011).

The RMSE formula can be formulated as follows:

$$RMSE = \text{sqrt}(\sum((X_0 - X_1)^2 + (Y_0 - Y_1)^2)/n)$$

Where X_0 is the delineated stream and X_1 is the PDOK reference dataset. Evenly distributed points along the line segments are created. The number of sample points is represented by *n.* The RMSE of the average distance of the points from the reference points is being calculated. Furthermore, the sinuosity is being calculated to examine the flow path of the stream to each other. The sinuosity is the ratio of stream length to valley length (Anderson & Ames, 2011). This can be calculated and compared for each stream with:

$$S = \frac{L_m}{L_s}$$

Where L_m is the meandering length (the length of the whole polygon of the stream) and L_s is the straight-line distance from the starting to endpoint of the stream.

Lastly, the vertical performance of the models is examined. By analysing the elevation of the streams per horizontal distance, the performance can be assessed locally. In the perfect situation, the longitudinal profile would be a smoothly descending line, without any upward spikes. Also, as the algorithm streams from point to point, the selected points of the algorithm are plotted in a graph to compare them to the vertical performance of the stream. Furthermore, the vertical performance of conventional D8 results and the performance of the absolute point-method are compared. The graphs are created with the Python library MatPlotLib and the terrain profile plug-in in QGIS. The script for these graphs can be found in Appendix B. Because a continuous surface model is required for an elevation graph, for both research areas a DTM is used for the longitudinal profiles.

A D8 stream delineation consists of different steps that must be conducted. These steps are shown in the workflow diagram in figure 12. An AHN 0,5m raster DTM serves as the input. Before use, the file was clipped to the extend, with some margins added, of the research area. The following steps were executed by using existing GDAL and SAGA libraries in QGIS (van der Kwast & Menke, 2019). First, no data cells are filled by interpolating from surrounding pixels. Then, to ensure the continuity of the flow, the sinks are filled using the earlier mentioned Wang and Liu (2006) algorithm, provided by SAGA. A low minimal slope of 0.01 degrees was used, because of the relatively low topographical relief, this low value was used. After completion, the data is prepared for the flow direction calculation to derive a channel network. This is done by an algorithm that calculates a flow direction raster based on the gradient value assigned to each pixel. The flows are classified from one to eight, as this method is limited to eight cardinal directions. Lastly, the streams are distracted from the flow direction raster with the Strahler order. With a process of trial and error of the ordering, the actual flows in the scope area can be identified.

Using these three validation methods, combined with visual interpretations, the accuracy of the D8 and the absolute point-method can be assessed. The result streams are visualised on the elevation point cloud points. To understand the topography of the area best, and to visualise a recognizable pattern in the gradient of the research area, the points are classified with an equal count classification. Furthermore, a grey-scale was used for visualisation of the elevation of the points to optimise the visibility of the streams in the maps. Using this greyscale shows the elevational pattern of the map and enables the viewer to identify all streams well. Because the prototype map consists of only one stream, a coloured ('reds') classification was used here to give an optimal insight into the height patterns of the surface.

4 Results

This chapter presents the results of the created prototypes. Based on the preliminary results of the prototypes, the absolute point-method was advanced. To retrieve the results for this research, the main focus was to make a completely working algorithm. This algorithm forms the fundament for an easily accessible processing algorithm in QGIS. The results and the presumable added value of the absolute point-method was lighted out. Also, the conventional D8-method is executed in the research areas. In the Olst research area, the topographical relief was too small to successfully apply the D8-algorithms. Therefore, this comparison is only made at the streams in the research area at Epen. First, the results of the prototypes are presented. Then, an overview of the result streams is given. Thereafter, the parameter optimisation process was discussed. The influence of different parameters will be discussed, based on the results of the algorithm with different parameters. Then, the quality of the different streams was discussed in detail, based on different quality parameters and visual interpretation. This chapter concludes with advice for the parameter settings for the absolute point-method.

4.1 Results of the prototypes

For this research three prototypes were developed. The results of the different prototypes are discussed in the following paragraphs.

4.1.1 TIN Network-based approach

This prototype was applied to both a normal Delaunay TIN as a constrained Delaunay TIN. A stream was manually delineated based on the slope, assigned to the edges of the TIN. Figure 18 illustrates that the stream does follow the reference stream and flows from uphill downwards roughly, but the delineation is not very accurate.

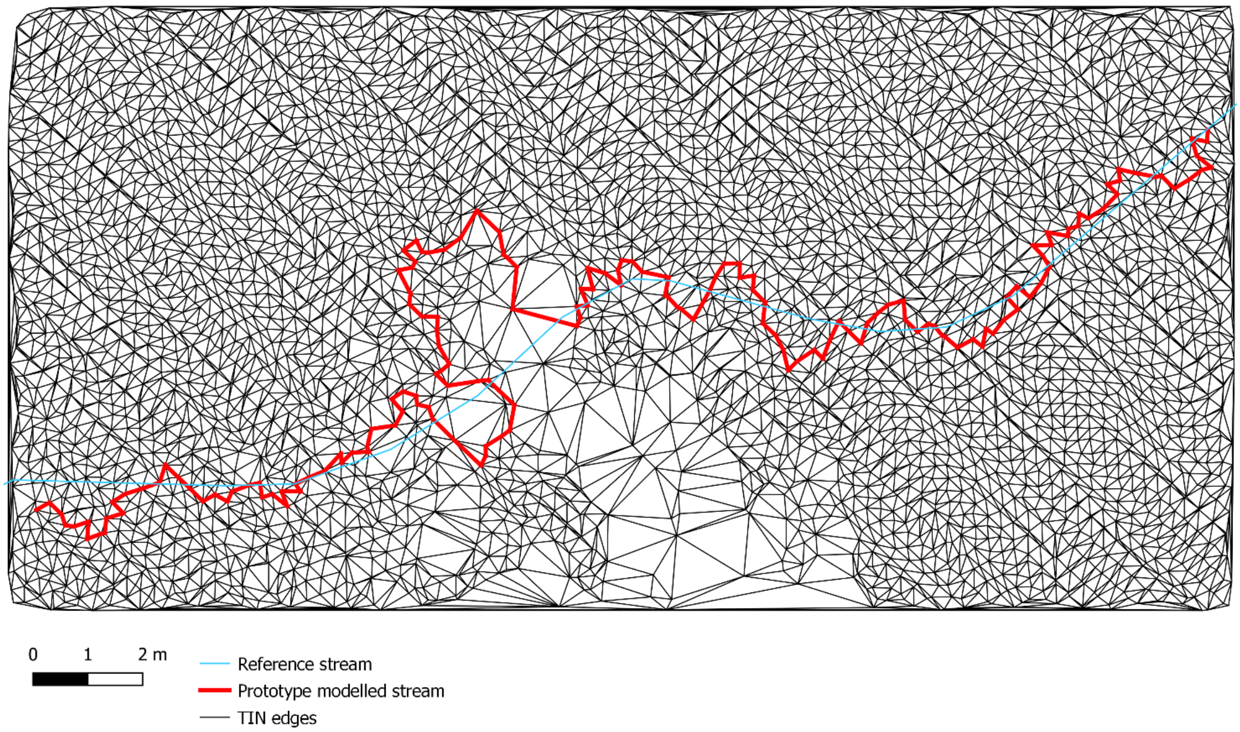


Figure 18: TIN network-based stream delineation prototype.

In the centre of the area, where there is a relatively low data density, the flow deviates strongly from the actual stream. Furthermore, the stream is very rugged and does not seem to be similar to a natural flow. This could be caused by the restriction for the algorithm to only flow over the edges. Hence, this algorithm is not satisfactory and does not meet the requirements for this research. Furthermore, a constrained Delaunay TIN was used as input for this method. However, it was not possible to accurately delineate a stream in this area, because artificial flat areas are created around the contour lines. This is shown in figure 19, where the edges without a slope are visualised in red. It can be concluded that the contour lines, that function as a constraint during the triangulation, cause flat areas.

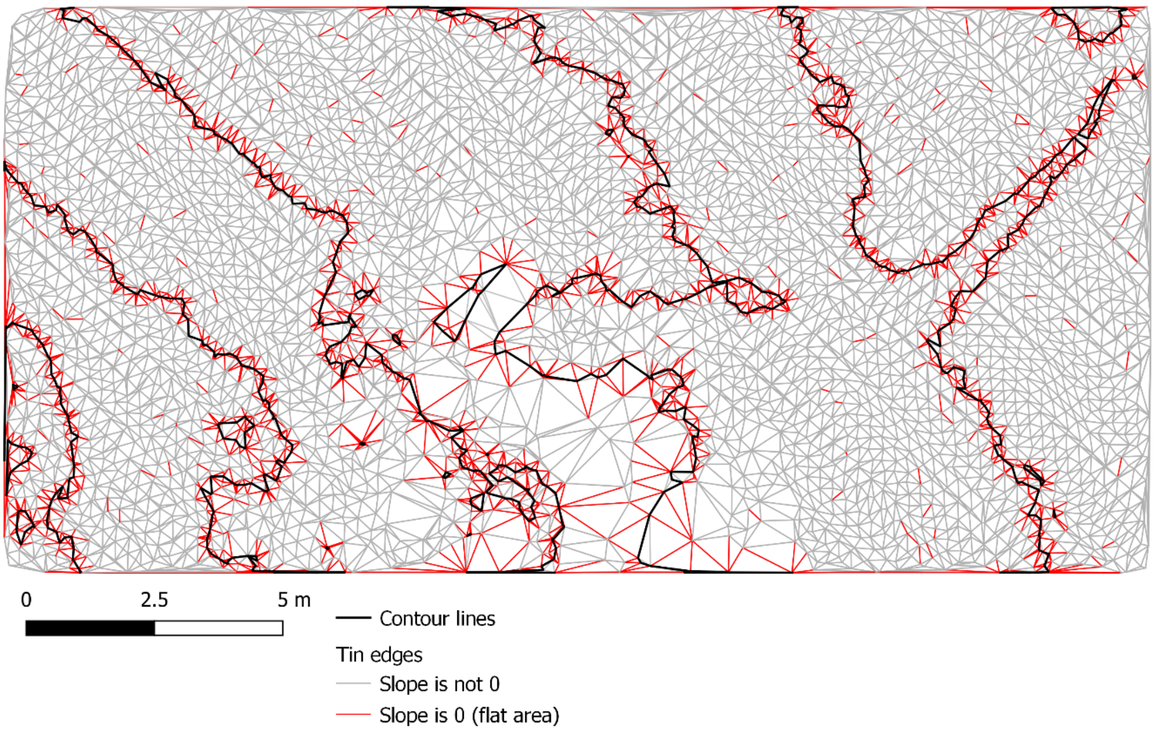


Figure 19: The red edges show that there are artificial flat areas created around the contour lines that are used as a constraint for the Delaunay triangulation of this dataset. These flat areas impede the model to calculate the stream downward.

A global approach, that derives a stream network for the whole area is possible with the TIN based approach. All edges have, based on the slope, a flow direction assigned. Therewith, a flow direction network can be identified. However, based on figure 19, the modelled streams will still not follow a natural path and illogical course changes will presumably occur. Moreover, a flow network from a Delaunay triangulation was extracted by Gabrisch (2011). This method worked and was relatively accessible as it was written in Python, but was effective in an area with a relatively high relief with elevation differences up to 150 meters. Although a global approach, where the stream directions on the whole surface is taken into account can be desirable. A higher accuracy can be achieved by letting the water flow over the faces from edge to edge. However, this method did not meet the requirement of simplicity. This can be concluded from the derived stream with the prototype, which is not accurate enough.

4.1.2 Normal vector approach

Another suggested approach is the normal vector-method. A prototype was developed for this, but many challenges where faced. The designed method follows the stream over the edges of a Delaunay TIN from a specific point uphill to the outlet. However, the normals led the path in random directions as the normals at the stream itself pointed towards each other. Therefore, a working prototype for this method is not feasible. Figure 20 illustrates that the orientation of the points at the stream is pointed at the centre of the stream. Although the points farther away do have an orientation in the direction of the stream, the normals on the actual stream are disrupting the model. This results in a back and forth going flow, which is not an accurate representation of the actual flow of the water. The sinuosity of the modelled flow would presumably be too high. Furthermore, as illustrated in figure 20, the gradient of the normal vectors becomes more significant as it comes closer to the reference stream. The bucket approach of Dakowicz and Gold (2007) was found too complex to implement in combination with the normal vectors. However, this could be an effective method if looking for a more complex, demanding and advanced algorithm.

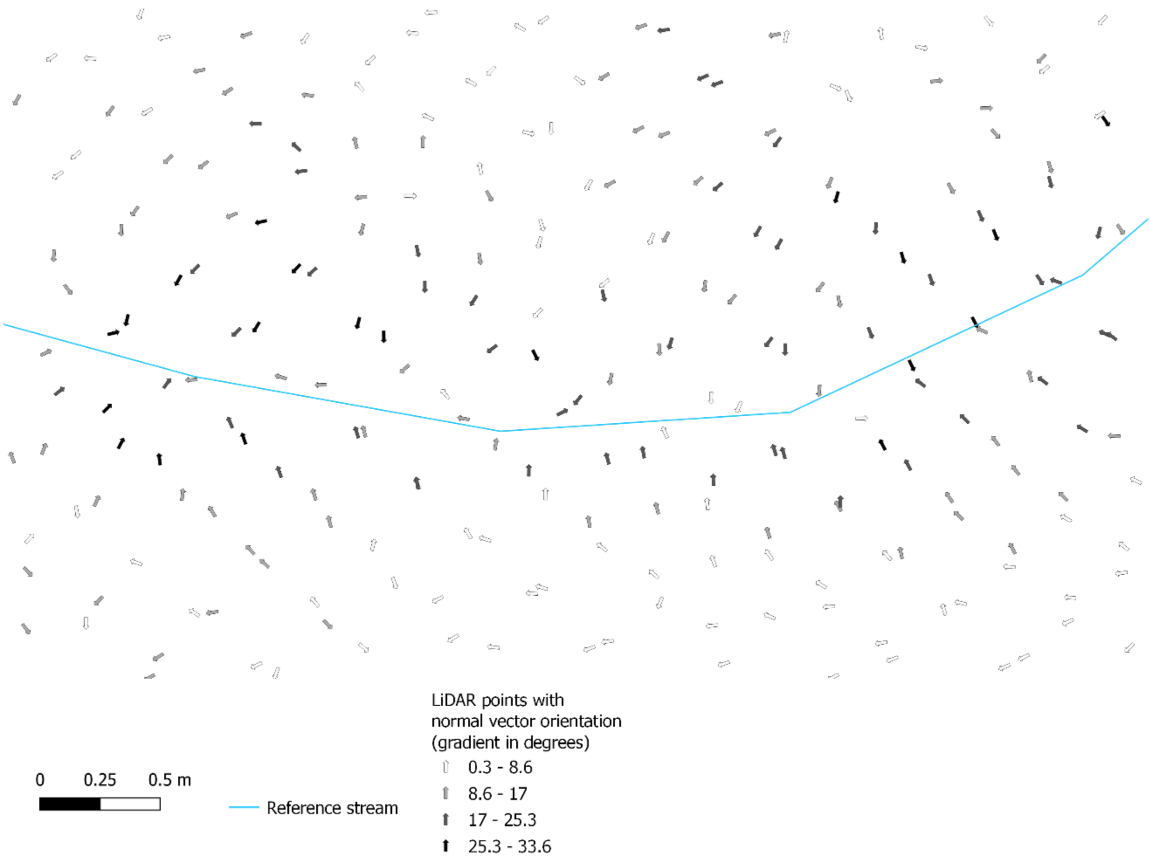


Figure 20: Normal vectors of the points show that the directions at the stream cause disruption.

4.1.3 Absolute point-based approach

Lastly, the absolute point-based method is suggested. This method selects the lowest point in a specified neighbourhood. On this relatively small prototype dataset, there were no performance issues. However, if the datasets become bigger, performance problems could occur. This can problem can be mitigated by limiting the required calculations. Hence, it is not necessary to calculate the lowest neighbour for all the points. It is only required for the points where the model streams to. Furthermore, this method is a very local approach as it looks at the individual points. It is possible that the lowest point that is identified within the search radius, is positioned in a local minimum. To circumvent this, the search radius can be extended locally to identify a lower point. A search radius of 2 meter was used in this prototype; the extended search radius was 3 meters. It can be observed in figure 21 that the extended search radius was applied in the centre of the stream in this prototype. A total of 15 line segments were created during the execution of this prototype and the length of the stream is approximately 25 meters. Lastly, it was also tested to delineate the stream from downhill upward. This did not work as the model flows to the edge of the research area. Logically, the points outside the stream bank are higher than in it. In figure 21 it is demonstrated that this method is relatively accurate by comparing the modelled stream to the reference stream of PDOK. Although this prototype is very basic and is executed in a small area, it was found to be suitable to scale up. Since the lowest neighbour for every point was calculated, this method seemed to be very demanding. This was taken into account during the creation of the advanced algorithm. In the advanced algorithm, only for the points that were used, the lowest neighbour was identified. Also, spatial indexes for the points were created to optimise the database speed.

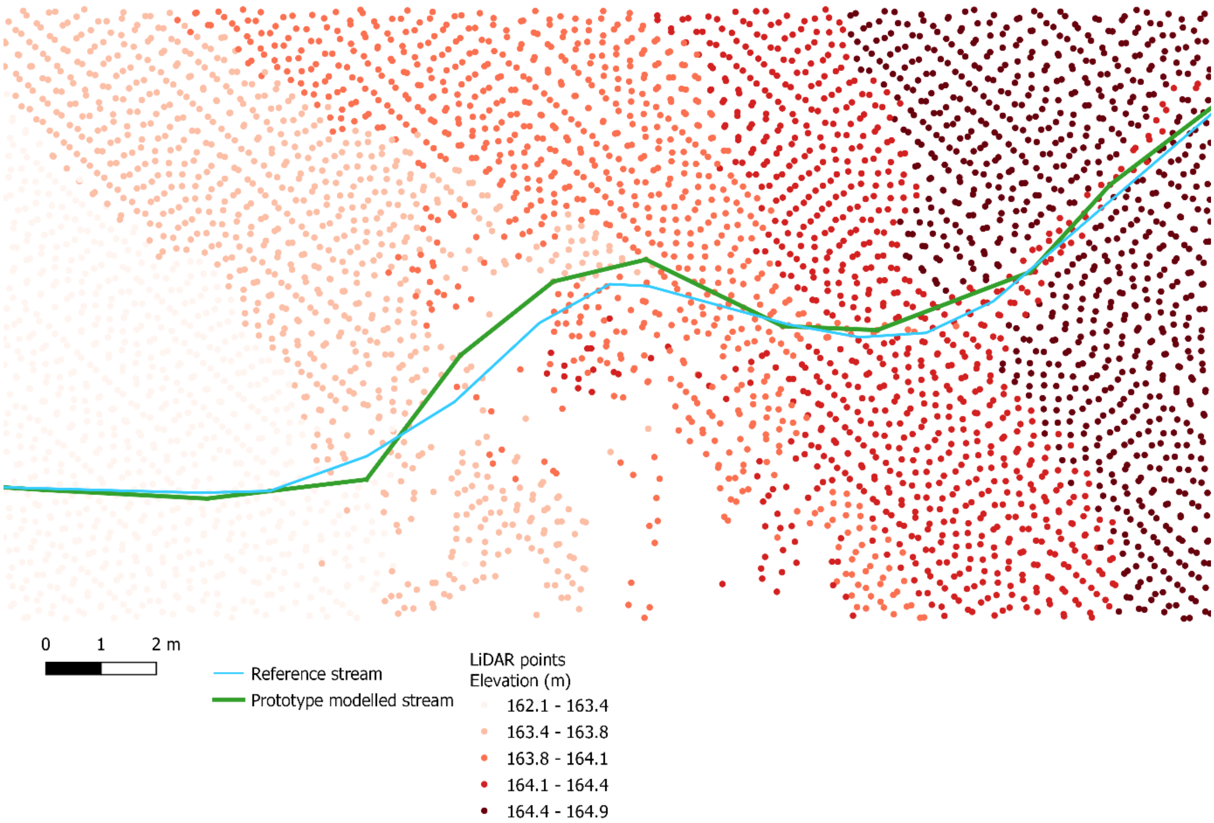


Figure 21: Lowest point within neighbourhood approach prototype.

This approach seems to meet the requirements that are set for this research, as this approach is accessible (all operations can be executed in QGIS), accurate (the modelled stream seems to follow the reference stream accurately) and simple (the processing capacity is not too demanding and by indexing the points, processing requirements can be improved). Although this method is limited to a single stream delineation, it is expected to be effective in low relief areas and offers a solid addition to retrieve information about water flows, where conventional methods are less accurate or return errors. The added value of this approach, compared to conventional methods, was further investigated in this research. The arbitrarily selected starting point can be selected uphill or could be selected at the location where the error of conventional methods begins.

This method is a very local approach, as the algorithm only looks at the stream direction per point. Where TIN and rasters provide a local stream direction for each location in the dataset, does this approach only looks at the stream direction from point to point. This, however, contributes to the simplicity of the model and the high accuracy of a single stream delineation can be achieved. It is expected that this method, with its higher accuracy combined with the relatively high simplicity, is of added value to the existing methods.

Finally, an answer to sub-question 2 can be provided: *"What (combination of) approaches with direct point cloud analysis is most suitable to delineate streams in an area with low topographical relief?".* The absolute point-method seems to meet the requirements for this research most closely. The integrity of the point cloud data is preserved and the natural flow of the water is modelled accurately. Because of the preservation of all the measured points and the high data density, the accuracy of the model is expectedly higher than conventional methods. Therefore, the reliability and accuracy of this method presumably is higher than other suggested methods. This method can be created with existing tools and libraries in (Py)QGIS. Therewith, the requirements of simplicity and accessibility are met sufficiently. Although this approach is very local, the algorithm handles with sinks by locally increasing the search radius if no lower point is identified and by pushing the stream forward implementing a maximum allowed course change. Stream delineation with Delaunay triangulations is also possible but does not sufficiently meet the requirements for an algorithm sought for this research. Given the low elevation differences, the stream would become too rugged and accuracy will be not sufficient. This can be improved by using a more advanced algorithm with a Delaunay TIN, but then, the requirement of simplicity and/or accessibility would not be met. The absolute point-method meets the requirements most closely of all the discussed methods.

4.2 General results

Based on the previous results, the absolute point-method was created. This method is used to gather the results and to prove the added value of an approach that looks more direct at the point cloud. As the main objective of this research is to show the added value of the absolute point-method, the absolute point-method is not yet fully functioning as a processing algorithm in QGIS.

Currently, the algorithm can only be executed from the Python-code (attached in appendix A). To ensure the open availability of this algorithm, the source code can be found on: https://github.com/stn228/Absolute_point_method_Stream_delineation. As this research focusses on the provision of a proof of concept of this method, a well-working, user-friendly tool has not been created. However, this code forms the fundament for a ready-to-use tool with a user-interface in QGIS. By making some minor changes, the tool could look like (and work) as is shown in figure 22 (*QGIS Documentation*, 2020). Conceptually, this algorithm works well, but to make it really user friendly, some changes could be made. The most important change is the automatic identification of the x, y and z coordinates. Also, the lowest point in the area must be identified automatically by the algorithm, or another condition for termination should be included. For instance, the calculations automatically end if the stream is within a specified distance from the edge of the research area.

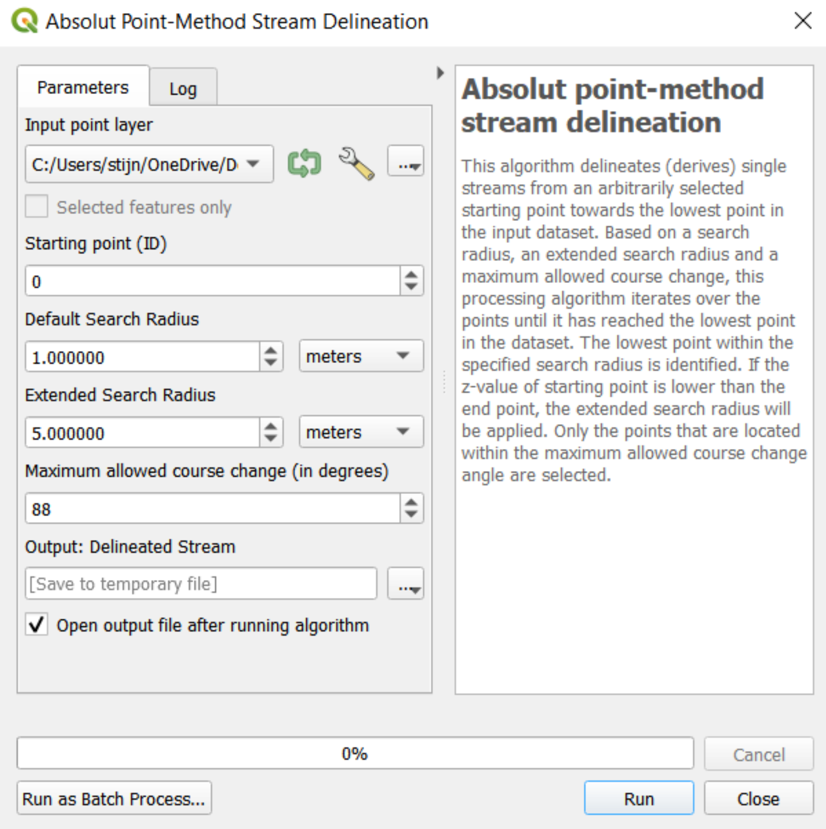


Figure 22: This figure shows how the algorithm can be adopted in QGIS as a ready-to-use-tool.

4.2.1 Overview of input data

For both research areas, a filled raster DTM and a point dataset was generated. This data is used input-data for the analyses. Table 3 gives an overview of the statistics of the different input datasets. The number of cells in both areas is much higher, but the difference is not as big as expected, as there are averagely more than five points per grid cell. However, the raster grids are clipped to a squared extent, while the point data area is smaller as it can be clipped to the exact area extent. The point dataset was minimized as much as possible to boost processing performance. In line with the theoretical expectations, the standard deviations of the DTMs are higher than the point cloud as the accuracy of the point cloud is expectedly higher. Also, the maximum and minimum values of the rasters are respectively higher and lower than the point data. Since the points, originally serve as input for the raster DTMs, this is remarkable. Since the squared IDW interpolation preserves the maximum and minimum values, this can have two causes: (1) the minimum and maximum value are outside the point dataset extent (as the raster is a bigger area) or/and (2) the Wang and Liu (2006) fill sinks algorithm changes values to smoothen the raster dataset to increase the accuracy performance of the D8-algorithm. In the case of Epen, where a difference of 13 meters can be identified, it is caused by the first listed reason. In the corner of the dataset, a higher area is located. This area is not included in the point dataset. The dataset was not clipped to the exact extent of the scope area as this could cause errors during the execution of the D8-method. If the raster is clipped to the geometry of the scope area, the height differences are significantly lower in Epen and Olst. The minimum maintains the same, but the maximum becomes respectively 190.2 and 8.3 meters.

Table 3

Statistics of input datasets for D8- and absolute point-method

Statistic	*Filled DTM Epen*	*LiDAR ground points Epen*	*Filled DTM Olst*	*LiDAR ground points Olst*
Size of area (m²)	54 400	36 527	141 246	75 737
Cell size/point density	0.5 m²	13.2	0.5m²	9.1
Number of cells/points	217 600	482 662	564 984	686 181
Mean	178.72	175.01	3.52	3.75
Std. deviation	9.24	6.52	1.51	0.49
Minimum	159.91	160.00	1.00	2.78
Maximum	202.75	189.57	8.32	8.19

4.2.2 Research area 1: Epen

In the Epen research area, three streams are modelled. Two streams from the same starting point, but the streams have disparate search radii. These parameters and the impact of the differences are discussed later in this chapter. The third stream starts from another starting point. This starting point was selected because the D8-modelled stream altered from the reference stream. The conventional D8-method resulted in a stream network in this research area. For the analysis, the streams that have the same starting point are selected. The results are shown in figure 23. This map is also attached in Appendix C. Based on the first visual impressions, both models perform satisfactorily in this area as they all flow from the starting point in a logical and natural path towards the outlet. Looking at the underlying data, this was expected, because a clear pattern from high to low elevation can be distinguished. A quantile classification is adopted for the visualization of the points.

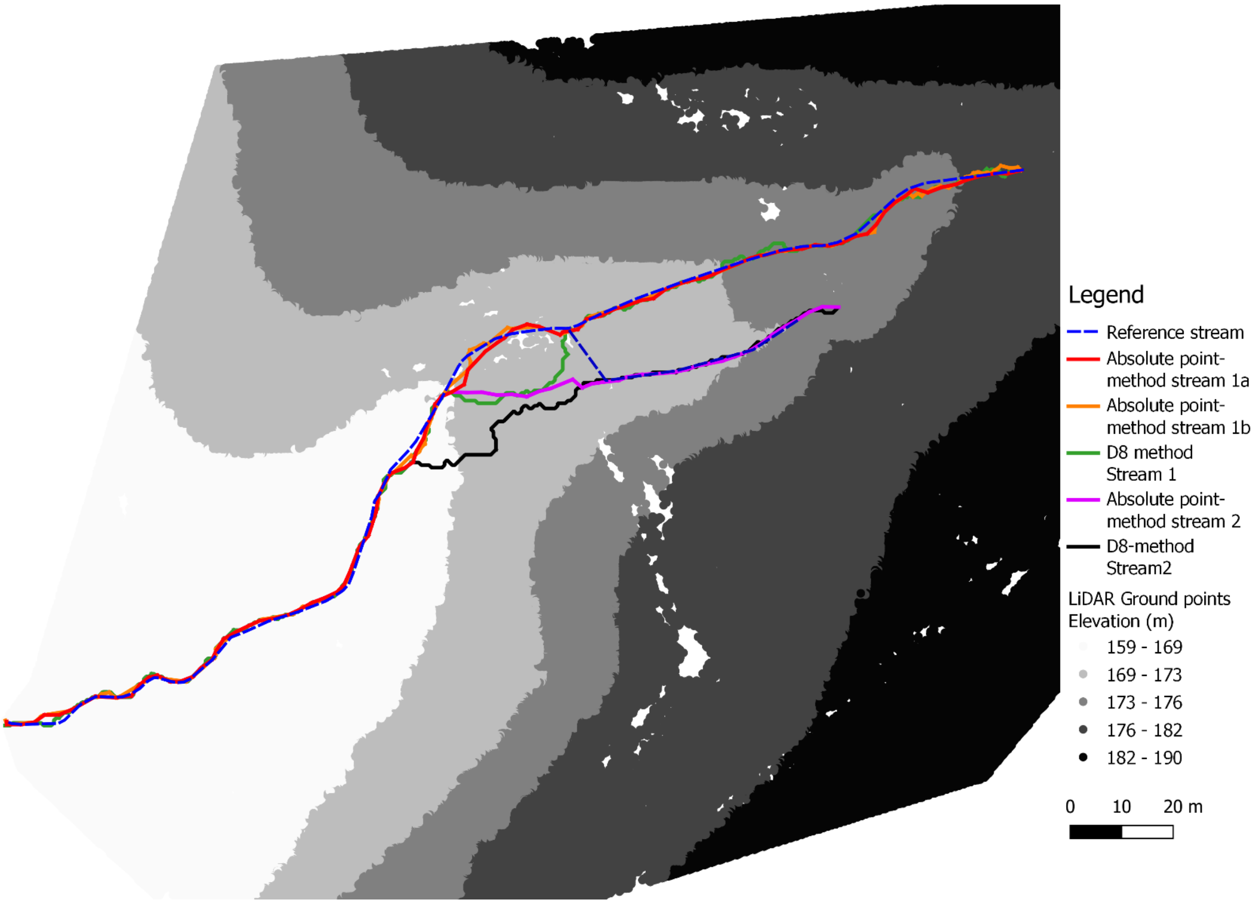


Figure 23: Overview of results in Epen.

4.2.3 Research area 2: Olst

For the research area in Olst, a less clear pattern from high to low elevation can be distinguished. As stated before, the runoff in this area is almost negligible. Therefore, stream modelling with the conventional D8-method was not feasible in this area. Execution of this method did not result in a stream network, because the flow directions differ too much. However, as is shown in figure 24, the absolute point-method did perform well at first glance. This is also attached in Appendix D. Because of the complex nature of this area to derive stream networks, careful parameter optimisation was required. This is discussed next in this chapter.

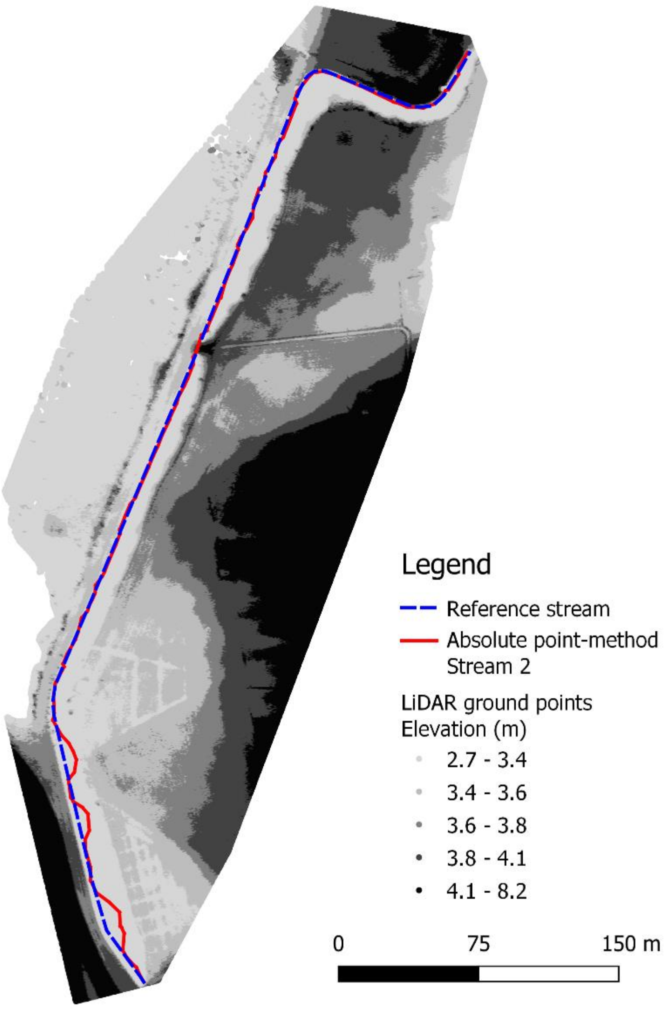


Figure 24: Overview of results Olst.

4.3 Parameter optimisation

As mentioned before, different parameter settings can influence the results of the algorithm strongly. Mostly the accuracy and the runtime is influenced by the chosen settings. Table 4 gives an overview of the chosen settings for the different modelled streams. The trade-off between model performance and accuracy can be seen in the results of stream 1a and 1b. Stream 1a is less accurate than 1b (table 5 and 6, page 61), but the runtime for the calculations is significantly higher. Stream 2 (Epen) has an alternative starting position and delineates another part of the reference stream. This is caused by the number of line segments that must be calculated. The line segments are the polylines that are created during each iteration of the algorithm. The shorter the average line segments, the more data points are included in the stream. This presumably increases the accuracy of the algorithm. If a high accuracy is preferred, the average length of the line segments should be minimised.

Table 4

Overview of parameter settings per stream

Stream	*Search radius (m)*	*Extended neighbourhood (m)*	*Maximum allowed course change (°)*	*Number of line segments*	*Average segment length (m)*	*Runtime (s)*
Epen 1a	3	5	+/- 88	97	2.54	264.11
Epen 1b	1	5	+/- 88	163	1.54	604.40
Epen 2	3	5	+/- 88	78	2.57	220.45
Olst 3	6	10	+/- 50	113	5.56	815.11

In the Epen region, a maximum allowed course change of +/- 88° was sufficient to ensure the stream flows forward. Also, the search radii can be set to particular low values. However, in Olst, with the extremely low topographical relief, the search radius must be bigger, otherwise the stream ended up in a local depression. Also, the maximum allowed course change had to be decreased. Figure 25 illustrates the stream with an 88° maximum allowed course change. Because all the values are very close to each other, the stream flows back towards the starting point. When comparing the runtimes, stream 3 underperforms per line segment. Stream 1b has more line segments and has a lower runtime. The streams in the Epen research area need between 2.7 and 3.7 seconds per line segment, while the model needs 7.2 seconds per line segment. This is caused because the dataset is slightly bigger, but mostly because the extended neighbourhood procedure had to be executed significantly more in Olst than in Epen. This takes longer as the algorithm only identifies a lower neighbour in the extended search radius if there is no lower neighbour found in the default search radius.

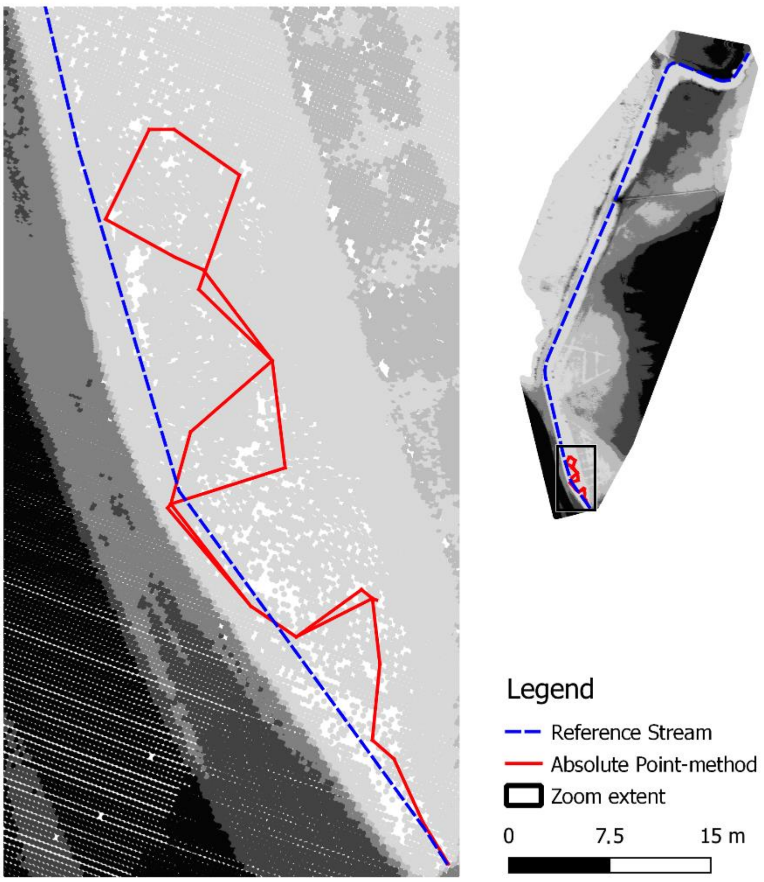


Figure 25: An example of the results with wrong parameter settings.

4.4 Accuracy

The accuracy of the streams is mainly examined with visual interpretation, the RMSE and the sinuosity, compared to the reference stream. As shown before, this research distinguishes four modelled streams. These are discussed stream by stream, except for stream 1a and 1b, which are discussed together.

4.4.1 Modelled stream 1a and 1b: Epen 1

The results in figure 26 shows that both stream 1a and 1b follow the reference stream accurately. This is confirmed by the low RMSE for both streams in tables 5 and 6. Also, the sinuosity of the streams, derived with the absolute point-method, matches the sinuosity more closely than the D8-stream. The relatively small default search radius of stream 1b results in a highly accurate stream with a very low RMSE. The RMSE of stream 1a is almost twenty times more accurate than the conventionally delineated stream. However, 1a also performs considerably better than the D8-derived streams. Furthermore, figure 25 shows that the D8-stream follows an alternative path at the centre of the stream. At this point, trees cause data voids and less LiDAR ground points are retrieved here.

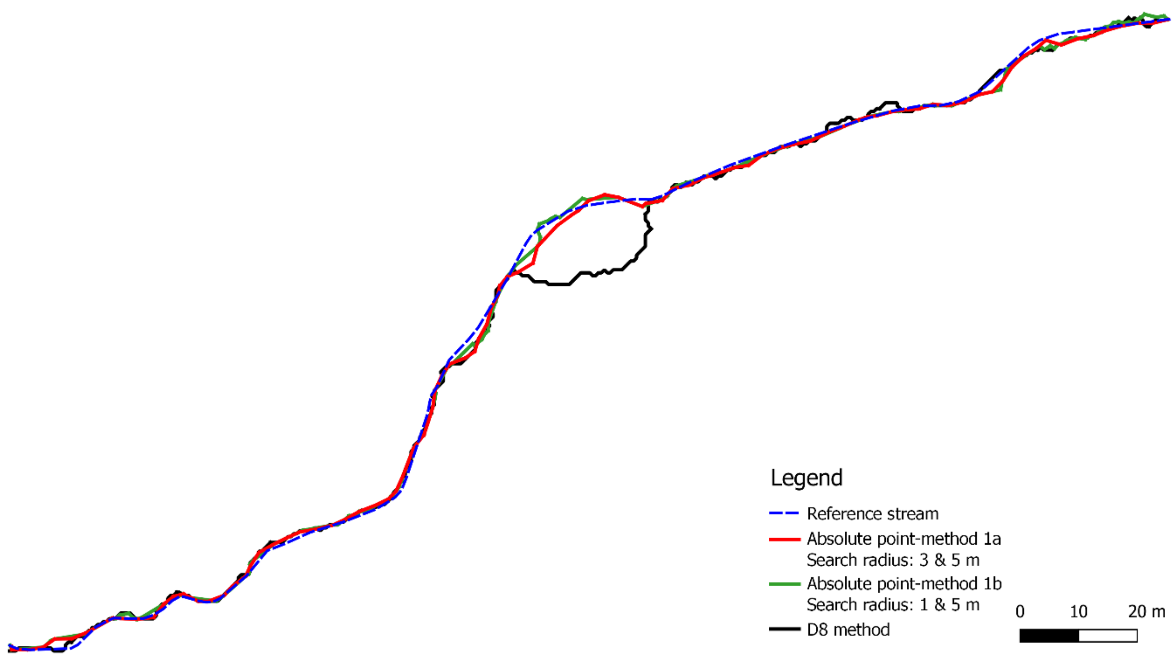


Figure 26: The reference stream and the D8-derived stream compared to modelled stream 1a and 1b.

Table 5

Accuracy of stream 1a (3m default and 5m extended search radius). The Euclidian distance of the path: 224.5 meters

Stream	*Length (m)*	*Sinuosity*	*RMSE (m)*
Reference stream	242.6	1.08	-
D8-method	276.3	1.23	10.32
Absolute point-method	246.3	1.10	2.06

Table 6

Accuracy of stream 1b (1m default and 5m extended search radius). The Euclidian distance of the path: 224.5 meters

Stream	*Length (m)*	*Sinuosity*	*RMSE (m)*
Reference stream	242.6	1.08	-
D8-method	276.3	1.23	10.32
Absolute point-method	251.2	1.12	0.56

Lastly, the vertical performance of the stream 1a, 1b and 2 is sufficiently good. As is shown in figure 27, a smooth path down is followed. The that are selected by the algorithm follows the elevation accurately. Some minor peaks are identified. This is caused by distortions in the data. The peaks in the elevational values occur at locations where less data is available, as these peeks are identified at the location where trees are located. Since the ground points are used, the data density at these points is lower and, thus, data distortions occur. These peaks do not show in the D8-derived stream, as this is based on interpolations, where these small peaking values are smoothened due to averaging algorithms. Only the vertical performance of stream 1a is given, while the graph of profile 1b and 2 is more or less the same and are given in Appendix E. Figure 28 shows that, although the absolute point-method results in a more rough path, the difference is trivial and the vertical performance for both streams is considered as sufficient.

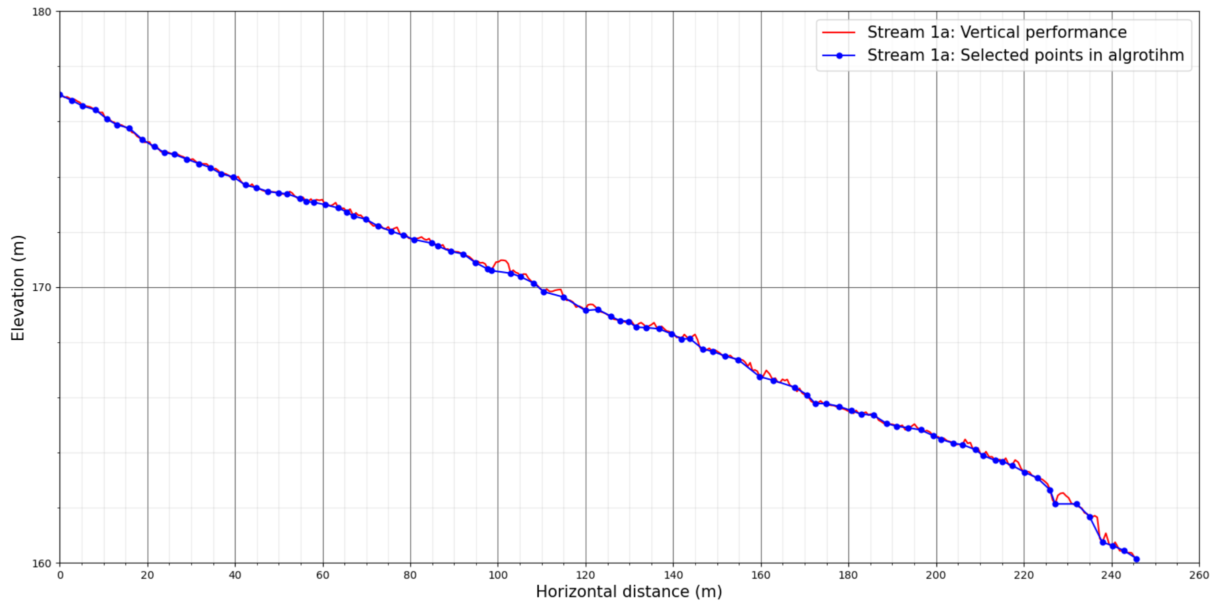


Figure 27: The elevation and the accessory horizontal distance of each point that is selected during the calculations for stream 1a. A smooth path downward can be recognized, without any upward stream line segments.

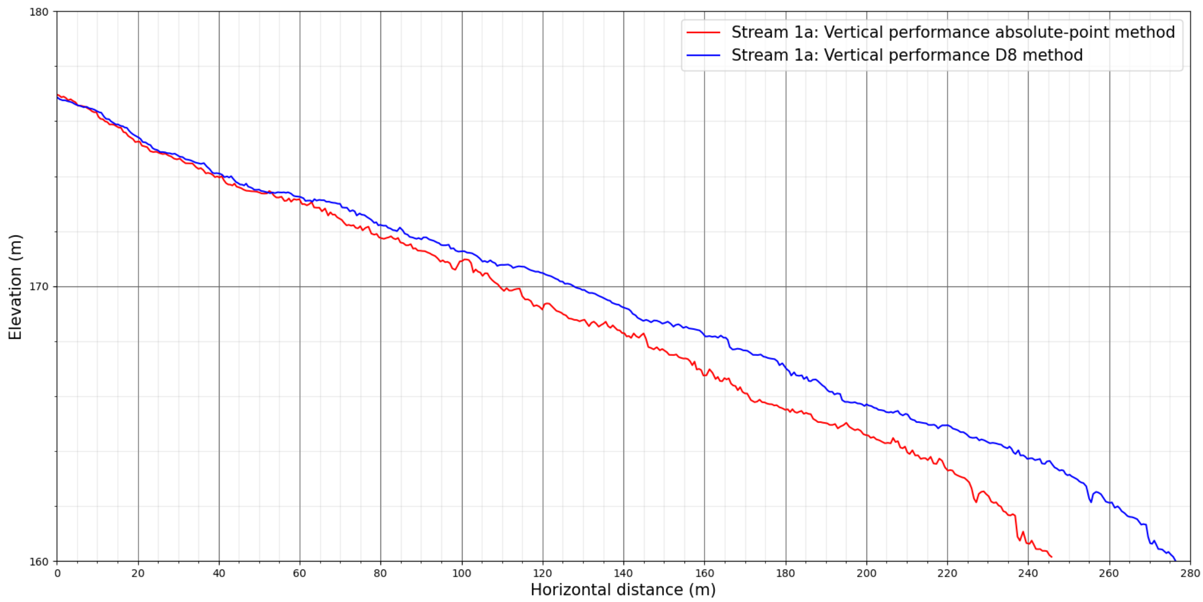


Figure 28: The vertical performance of the D8- and absolute point-derived streams compared for stream 1a.

4.4.2 Modelled stream 2: Epen 2

Figure 29 depicts the three streams from another starting point in the Epen research area. Here, all three streams follow an alternative path. Therefore, the RMSE of these streams is significantly higher than with streams 1a and 1b. Although the RMSE is much higher than at stream 1a and 1b, the absolute point-derived stream joins the path of the reference stream earlier, thus, the RMSE is lower.

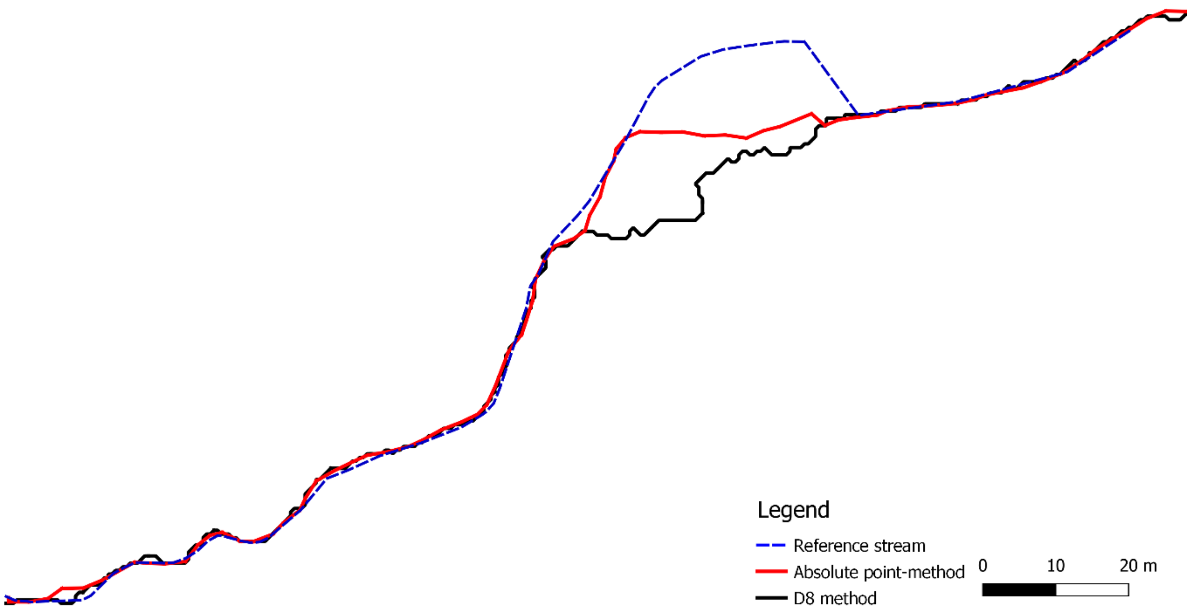


Figure 29: The reference stream and the D8-derived stream compared to the modelled stream 2.

Table 7 shows that the D8-stream is significantly longer than the other two streams. This is also the case at previously discussed streams. This is not only caused by a longer path that is chosen by the algorithm, but it is also caused by the cascading path that is caused by the restriction to stream to only eight directions. With that, the sinuosity of the D8-streams matches the sinuosity of the reference stream less good. Figure 30 on the next page zooms in on the different streams more closely. It confirms the hypothesis that the D8-streams are cascading and that the absolute point-stream follows a more natural flow path. This figure also shows a systematic difference between the reference stream and the modelled streams of approximately 0.5 meters. In this case, the modelled streams are possibly more accurate than the reference stream as they both indicate the same difference. Although the direction of the stream in both cases is accurate, the results of the D8-streams are limited by their restriction to only stream in eight directions. Although a good path is returned by the D8-method, this is not the optimal path (van Bemmelen et al., 1993). However, the comparison might be unfair as individual point data are compared to a 0.5m² raster DTM. Furthermore, a running average can be taken on the cascading D8-route to smoothen the path. In the case in figure 30, it would lead to results similar to the absolute point-method. Nevertheless, as is shown in figure 7, the overall results of the absolute point-method are significantly higher than the D8-derived stream.

Table 7

Accuracy of stream 2. The Euclidian distance of the path: 172.7 meters

Stream	*Length (m)*	*Sinuosity*	*RMSE (m)*
Reference stream	199.2	1.15	-
D8-method	220.8	1.28	23.01
Absolute point-method	200.7	1.16	11.8

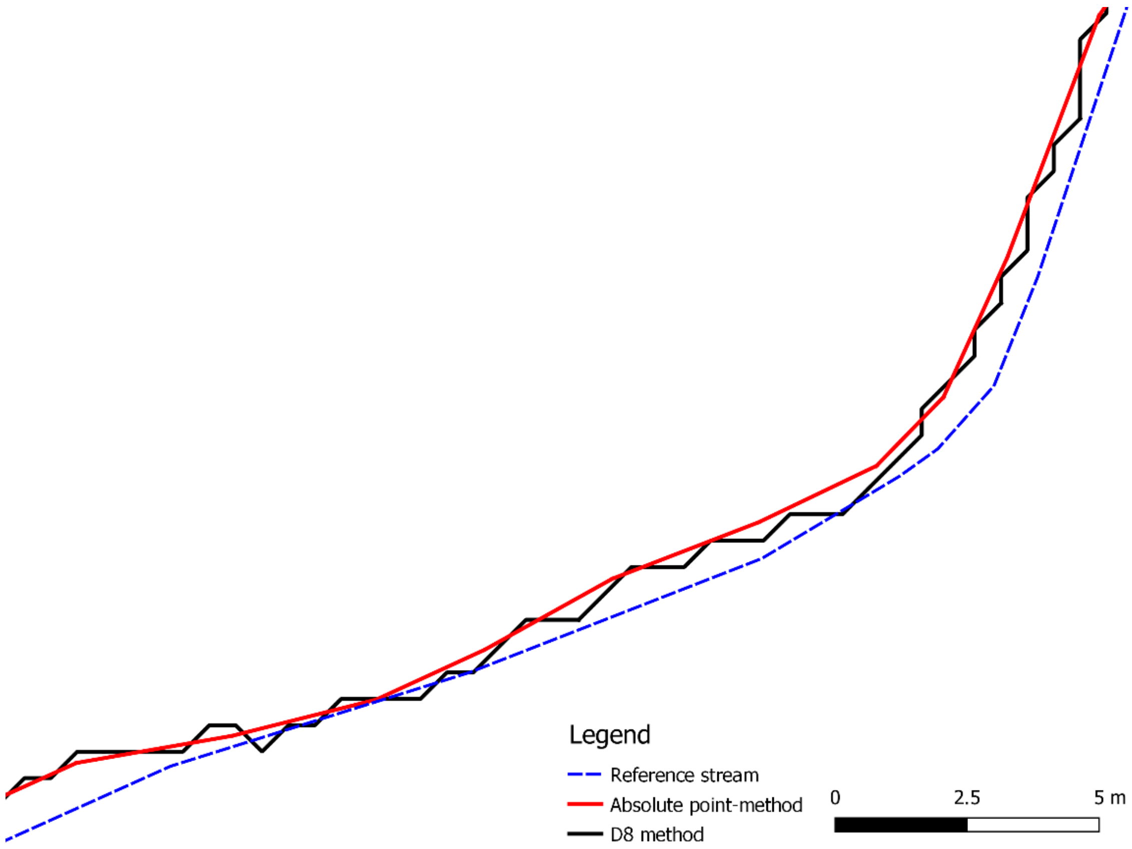


Figure 30: The D8-derived stream shows a clear cascading pattern while the absolute point-stream shows a more natural flow path.

4.4.3 Modelled stream 3: Olst

Finally, the third stream is modelled in an area with very low topographical relief. This stream flows from the south towards the north. The D8-algorithm was executable in this research area because the topographical relief was too low. The creation of the flow direction raster consisted of seemingly random direction in the cells and it was not possible to identify stream orders based on these results. The Strahler order resulted in streams of a maximum of three cells. This misfunctioning of the D8-method was expected (Gortzak et al., 2020). Although the D8-method was not operable in this area and the search default and extended search radius is relatively high, the results of the absolute point-method seem to be satisfactory. Figure 31 shows that the reference stream is followed by the absolute point-stream sufficiently. Furthermore, table 8 shows that the sinuosity is close to the sinuosity of the reference stream. The higher length of the absolute point-stream is caused by the divergence of the stream at the beginning, in the southern part of the stream. Except for this part, the reference stream is followed narrowly, also at the curves of the stream.

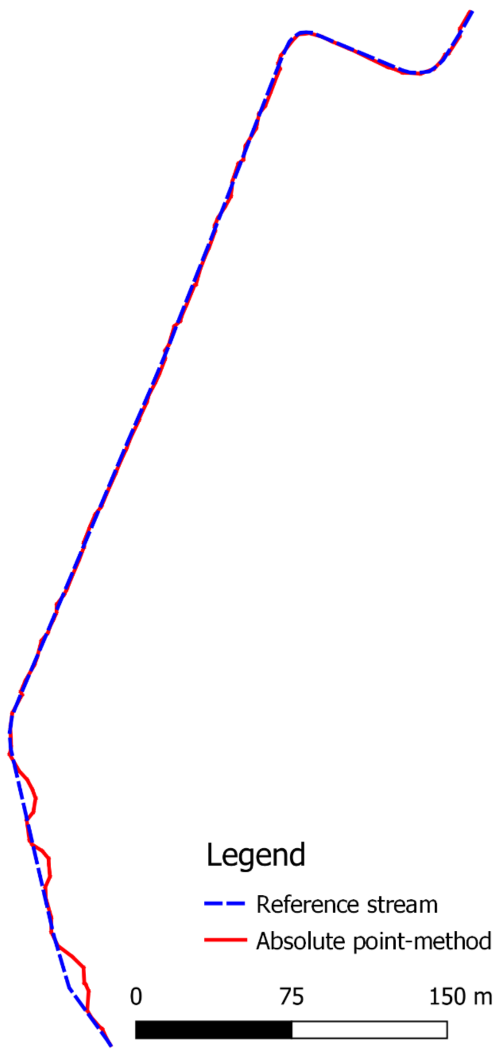


Figure 31: The reference stream compared to the modelled stream 3.

Table 8

Accuracy of stream 3. The Euclidian distance of the path: 526.71 meters

Stream	*Length*	*Sinuosity*	*RMSE (m)*
Reference stream	627.73	1.19	-
D8-method	n/a	n/a	n/a
Absolute point-method	644.12	1.22	8.54

Finally, the vertical performance of stream 3 was analysed. Figure 32 shows radical upward spikes in the longitudinal profile with almost not an identifiable path downward. This pattern was expected as this is similar to the longitudinal profile, discussed in chapter 3. This is caused by the sizable search radii and because this algorithm only looks at the start and the endpoint of the line segment. This allows the algorithm to pass areas with high elevation values. This is illustrated in figure 33, where a passage can be identified. The algorithm is not hindered by this, because of the high search radius that is applied. Thereby, this graph shows that the delineated stream can flow upward, against laws of nature. These laws are ignored to handle data inconsistencies and to overcome the very low topographical relief. Although it seems counterintuitive, it results in a relatively simple method to delineate streams in areas with such low elevational differences. In this case, the stream still is accurate because a culvert is located here. However, this could cause inconsistencies in the model. This is further explained at the end of this paragraph. The stream does follow the path towards the outlet. The line segments are smaller at the corners of the water, which is an important characteristic to follow the stream there accurately. In the straight parts of the stream, the line segments are bigger and therefore the longitudinal profile is very spikey. During the parameter optimisation, it is important to take this characteristic of the algorithm into account. When using a big search radius, the stream should be examined prudently to check if it crosses high elevation areas.

Furthermore, figure 33 shows data gaps at the presumable location of the actual stream. At some locations in the dataset, there is less data available if there was water during the data gathering since this research uses the ground points and these points are classified as 'water'. This results in the stream bouncing back-and-forth between the stream banks, where data points are present. Optimally, the model would follow the centre of the stream. However, this inaccuracy is relatively small and does not cause significant deviations in the model.

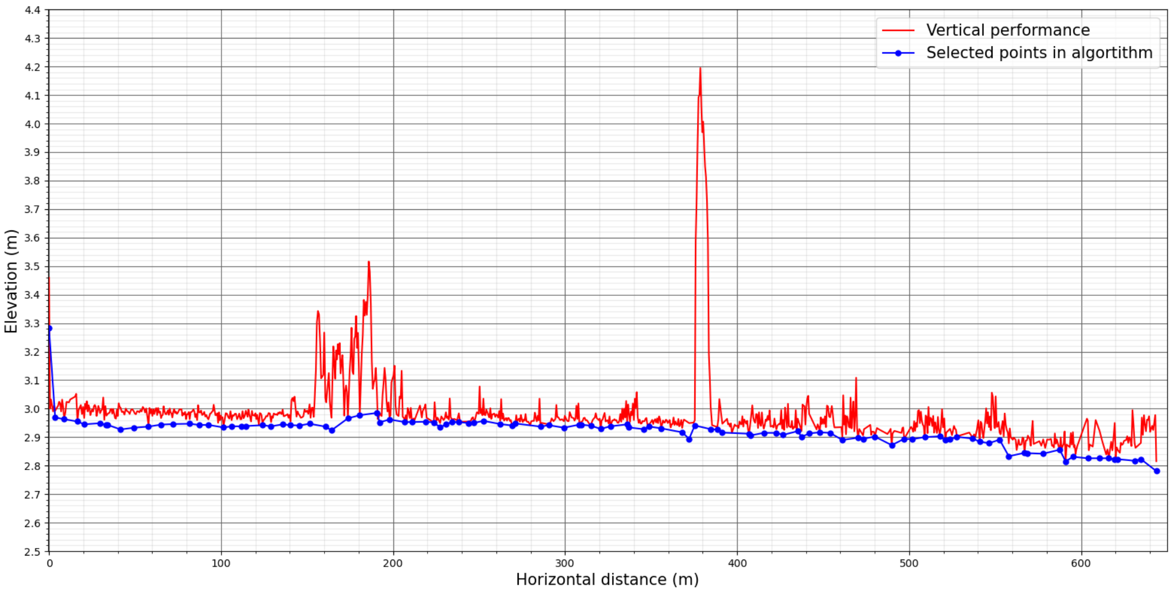


Figure 32: The elevation and the accessory horizontal distance of each point that is selected during the calculations. The path does not flow down smoothly and some line segments are streaming upward. This is caused by the very small elevation differences in the

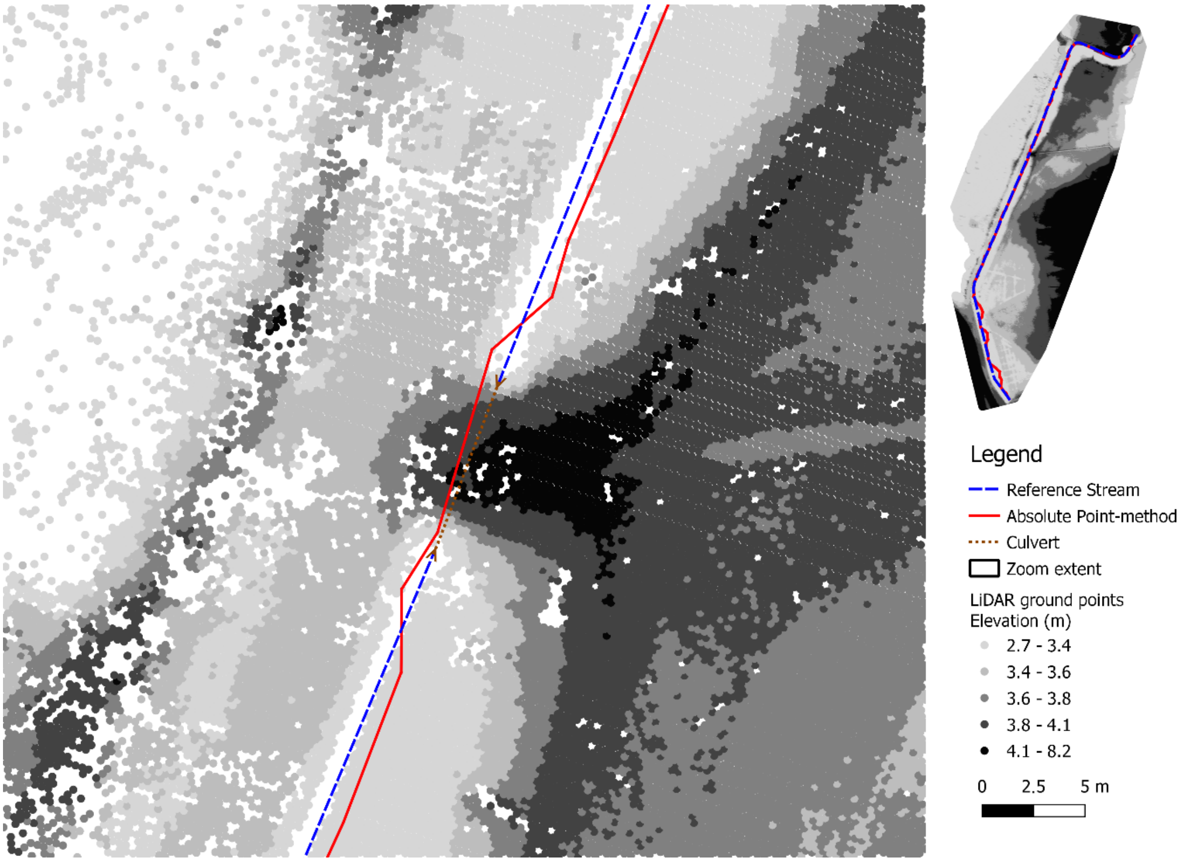


Figure 33: The absolute point-method enables the stream to cross areas with peaking elevation values; furthermore, the stream bounces from stream bank to stream bank.

As stated above, the vertical performance of the stream is very spikey. There are many high peaks. Therefore, the elevation of the 113 selected points in the algorithm were analysed. This is visualised in a graph in figure 31. The algorithm executes the extended neighbourhood function if the newly selected point is higher than the starting point. If there is also not a lower point in the extended search radius, the model draws the stream towards the lowest point in the search radius. Thus, the stream can flow up. As the effective elevation difference for this research area is about 0.2 meters, it occurred that the stream had to flow up. The maximum allowed course change predominantly makes the forces the stream to flow in the correct direction. The water flowing upward has two main causes: (1) because of the canalisation of this water stream, human interference impacts the natural flow of the water and (2) despite the high-quality data, it could cause inaccuracies in this area as there are parts in the dataset where there is no data, specifically at locations where the water is located (see also figure 32). The lack of data here is caused by the disability of the adopted LiDAR techniques to measure underwater effectively. Although this might be a strange pattern, the accuracy of this stream is still very satisfying. Mostly when comparing it to the D8-method, that did not return results because of this reason. Therefore, the ability of the stream to flow up can be considered as necessary, otherwise the algorithm will not return any good results. As is demonstrated in figure 24 (p. 50), if there are significant elevation differences, and if the parameters are set correctly, the stream only flows downward. The extension of the (extended) search radius for this stream did not lead to better results. Minimal changes in these parameters did not leat to another pattern in the horizontal accuracy. If the extended search radius becomes too big, the stream does flow downward only, but the accuracy of the results are strongly lacking then. Then, the model does cuts of all the curves of the original stream. Figure 34 shows that the channelised stream can be identified, based on satellite imagery. Also, the passage, where the culvert is located, can be seen. The other peaks in the longitudinal profile are presumably caused by overhanging canopy of trees. Based on this image, it can be concluded that the model follows the stream accurately.

Figure 34: A detailed aerial sattellite image of the Olst research area. At the location where the elevation peeks, a passage for presumably agricultural vehicles can be identified.

4.5 User recommendations for parameter optimisation

As stated before, this research provides recommendations to the user for parameter optimisation, based on the results. Although the results are strongly dependent on the characteristics of the chosen research area, some recommendations can be given. In this research, the Epen research area has an average declining slope of 0.07 degrees. Based on the longitudinal profile of this research area, a smooth decline can be identified. Stream 1b shows that, as the search radius decreases, the accuracy rises significantly. However, the calculation times increase as well if the search radius decreases. Also, the size of the research area, therewith the size of the input dataset, impacts the speed of this algorithm. The trade-off between speed and accuracy needs to be considered by the researcher. A considerable improvement of accuracy can be found with a default search radius of 3 meters and an extended search radius of 5 meters. If a higher accuracy is desired, the default search radius can be lowered to one meter. Also, the extended search radius can be made smaller. However, the default search radius for these areas should be between 1 meter and 5 meter to retrieve satisfying results. The extended search radius should be between 2 and 5 meters. The maximum allowed course change can be set to the default of 88 degrees.

For the Olst research area, a declining average slope of 0.0008 is identified. To handle the very low elevational differences in this area, the search radius must be bigger. This very small topographical relief results in many local depressions. Therefore, it is advised to have a search radius between 5 and 8 meters and an extended search radius between 8 and 15 meters. It is advised to lower the maximum allowed course change to 50 degrees to avoid the stream from circling.

It is important to note that the results and the parameter optimisation is strongly dependent on the characteristics of the research area and the input data. Therefore, the user should carefully examine whether the results make are logical or not. This can be done with visual interpretation, but also by looking at the vertical performance and the sinuosity of the streams.

5 Conclusion

The main objective of this research was to answer the following research question:

"What is the best approach to delineate streams in areas with low topographical relief using LiDAR retrieved point cloud data?"

An answer can be given by analysing the results of the formulated sub-questions step-by-step. The first sub-question resulted in an elaborate literature review:

1. *"What different methods for stream delineation with LiDAR data exist and what strengths and weaknesses can be identified?"*

Different data preparation, conversion and processing methods were discussed. The main advantage of direct point cloud analysis is that very high-quality data is optimally utilised. Other approaches lack in optimally using the high-quality data because they are based on interpolations and generalisations of the data. The weakness of point cloud data is the size of the datasets which result in demanding analyses. Therefore, if the data is not used and analysed efficiently, the methods can be complex and poorly performing. Therefore, three requirements for the new method were drawn up: accessibility, simplicity and accuracy. This means, respectively that the new method should be openly available for everyone, is more accurate than conventional approaches and is relatively easy to apply. Based on these requirements, three candidate methods were identified that are considered as most suitable that meet these requirements and directly use the LiDAR point cloud data. Three candidate methods were selected: a method based on a triangulation of the point cloud, a method using the normal vectors of the points and a method that uses the absolute (non-changed) elevation values directly from the point cloud.

2. *"What approach with direct point cloud analysis is most suitable to delineate streams in an area with low topographical relief?"*

The main aim was to find a method that accurately delineates streams, is simple to apply and is openly accessible. Prototypes were built for the three candidate methods. The preliminary results show that the absolute point-method meets the set requirements of this research best. Therefore, based on the theory and the results of the prototypes, this method was selected as most suitable for stream delineation in areas with low topographical relief. The absolute point-method looks for the lowest point within a specified search radius. To avoid getting in a local depression, an extendable neighbourhood and a maximum allowed course change was included in the algorithm. This maximum allowed course change results in that the search radius is being reduced to a half-circle or smaller. This process is iterated from point-to-point until it the outlet point is reached.

3. *"To what types of areas could a stream delineation with direct point cloud analysis be of added value?"*

It was decided to select two research areas. A distinguishment was made between an area with low relief and almost negligible relief with a decline of 0.2 meters In the first area, the conventional D8-method, which was used for comparison of the algorithms, was presumed to work better. This enables a more effective comparison of the results of the conventional and the newly suggested method. In the area with a very low topographical relief, it was expected that the D8-method would not return accurate results. To test this hypothesis, and to answer the sub-questions four and five, the absolute point-method was worked out more extensively. Therewith, it was possible to run the algorithm multiple times in different research areas.

4. *"Can an algorithm be created and applied in a suitable research area of the identified most suitable stream delineation method?"*

The designed prototype is a Python script that runs in the open-source software QGIS (Appendix A). In this phase, the absolute point-method was designed. This method looks for the lowest absolute point in a specified search radius, if no lower point is identified, the search radius extends. The stream is pushed forward by including an adjustable maximum allowed course change. By adjusting the parameters optimally, the algorithm worked for the research areas.

5. *"What is the added value of the selected approach in contrast to conventional methods to delineate streams?"*

After optimisation of the algorithm and execution of the conventional approach, the results show that the conventional D8-method did not return any results in the area with negligible relief; the absolute point-method, however, did. Additionally, based on the sinuosity and the Root Mean Square Error (RMSE), the accuracy of the absolute point-method was significantly higher than the results of the D8-derived streams in the Epen research area. In some cases, the absolute point-method resulted in almost twenty times lower than RMSE the conventional-method results. Careful parameter optimisation is very important for this method to achieve the desired accuracy and calculation speed. This is a trade-off, where the researcher can weigh the importance of calculation speed versus accuracy based on the size of the search radius. If parameters are not optimised satisfactory, the results can become very disturbed and incorrect. These disturbances can be identified by visual interpretation and by careful analysis of the longitudinal profile of the streams. This research gives guidelines for hydrologists who want to adopt this method. It must be noted that these guidelines are a broad approach and can differ per situation, as this strongly depends on the characteristics of the research area and the input data.

The vertical performance of the D8-method seemed to be slightly better in the Epen research area, as the absolute point-method shows a very spikey longitudinal profile. The vertical performance in the Epen research area was adequate for all models as a clear stream downward was identified, without any upward spikes. The vertical performance in the Olst research area was found to be very spikey and does not follow a clear path downward. This is caused by the extremely small elevation differences, possible human interference and some data gaps. Also, a smoother, more natural, path is followed by the absolute point-based method, as the D8-method forces the path into a cascading route, due to its restricted number of possible flow directions.

Mostly when looking at the RMSE and the sinuosity of the streams, it can be concluded that the accuracy of the newly designed method is strongly improved in contrast to the conventional approach. Therefore, the absolute point-method can be very useful and of added value in areas with very low relief, where conventional methods do not work or are less accurate. Since both conventional, as the new method, are accessible in open-source software, the accessibility is comparable. The absolute point-method offers a great alternative for stream delineations in areas with very low topographical relief as all tested streams perform better than the conventional D8-method. It can be used as a replacement if the D8-method does not work, or as an addition to the D8-method if the accuracy seems to be unsatisfactory. Thus, the added value of the absolute point-method, compared to other methods, is that it provides a significantly higher accuracy and offers decent simplicity. Furthermore, similarly to the conventional method, this method is openly accessible in open-source software. The Python- code for this algorithm is available on Github via: https://github.com/stn228/Absolute_point_method_Stream_delineation. As this method works in an area with practically the lowest possible elevational differences, the added value of this method has been proven. Thus, concludingly, the absolute point-method meets the set requirements of accuracy, simplicity and accessibility to delineate streams in areas with low topographical relief. As expected, this method profits from maintaining the integrity of the original data points by directly analysing the LiDAR point cloud. When looking for a method to delineate streams with direct point cloud analysis with the set requirements of simplicity, accessibility and accuracy, this absolute point-method can be considered as the, currently known, best way to delineate streams in areas with (very) low topographical relief with LiDAR data.

6 Discussion and future research

Where other methods lack or return errors, this algorithm has the power to accurately delineate streams directly with point cloud data. This research underlines the findings of different researches that emphasize the great potential of highly accurate analysis with point cloud data for hydrological applications. Especially Gabrisch (2011) and Anderson and Ames (2011) demonstrated the power of direct point cloud analysis. This research confirms their findings and suggests a method that is not only accurate but also is accessible and simply applicable for every hydrologist. By creating an algorithm that does not require rasterization of the data, several advantages appear. First, the high-quality LiDAR data is optimally utilized, as no data integrity-harming interpolations have to be done. The data is not generalised with averages of multiple points into raster cells. Furthermore, in contrast to conventional raster-based methods, this algorithm is not restricted to a limited number of flow directions, resulting in an unnatural flow path (van Bemmelen et al., 1993). Moreover, this method handles data gaps and sinks innovatively by including an adjustable search radius and maximum allowed course change. However, in addition to these measures, a fill sinks, data smoothening algorithm could also be created for direct point clouds. Similar to existing methods, this algorithm identifies sinks and raise the elevational value if all surrounding points are higher than the point of reference (Rheinwalt et al., 2019; L. Wang & Liu, 2006). Furthermore, the K-nearest neighbour algorithm could be used to comply with a minimum number of points within the search radius. Then, the user of the stream delineation tool could specify a number of points that should be within the search radius and the algorithm will adaptively change the search radius, based on the input data at a specific point. These features can be included in this method in future researches.

Counterintuitively, this research allows the stream to flow upward if there is no lower point within the extended search radius. The results show that this does not happen if the research area has a noticeable topographical relief. However, the stream does flow upward multiple times in the Olst research area. Looking at the longitudinal profile of this area, this is not unexpected. This unnatural pattern is possibly caused by human modifications of this area. The algorithm still returns satisfactory results because of the possibility to adopt an extended neighbourhood, but also because of the maximum allowed course change. This proves that this algorithm works in the area with unnatural patterns with the lowest possible relief. Although the water streams upward, the model still returns highly accurate results. By implementing this technique, the sink-problem is handled in a convenient and relatively simplistic manner. Because of the satisfactory results in the Olst research area, with almost negligible relief, it is assumed that this method works in almost all types of areas and could be of an added value to the existing methods. Although the results show that this algorithm fundamentally works, the test case areas are relatively small and the number of used points is not very significant for point cloud data. To strengthen the findings, and to improve the recommendations for the parameter settings, this method should be tested in bigger research areas, with more points and with different topographical characteristics. This will prove the scalability of this method and it will give more insight into the general performance of this algorithm.

Although the results of this research convincingly demonstrate the added value of the absolute point-method, some limitations of this method can be identified. First, this method does not take the volume of water streams into account. The significance of streams, based on cubic meters of water, is an important aspect to know. This would also help to derive a whole stream network in a specific area. This will enable the researcher to generalise small, insignificant streams away and to identify the more important streams. Thereby, this method is a very local approach. With this approach, it is only possible to identify single streams from an arbitrarily selected starting point. Therefore, it is not possible to extract the stream network of a specific catchment, based on the outlet-point. However, this often is desired in hydrological analysis. To extract a complete stream network, conventional methods must be adopted. Based on these results, the absolute point-method can be deployed to verify the accuracy of the delineated streams by the conventional methods. In future research to stream delineation methods with direct point cloud analysis, a more global approach could be investigated. For instance, the bucket-approach combined with the normal vectors of the individual points is potentially a global and effective method. However, the currently known tools do not allow to create such a method as the processing demands will be too high (Dakowicz & Gold, 2007). Also, a more advanced triangulation-based method could be adopted where the water flows in the direction of steepest descent from edge to edge, over the face of the triangle. This research identified a method where the water flows over the edges of the TIN. It was found that this is not an accurate approach and a better approach could have been selected. A global approach, that meets the set requirements for this research of accuracy, simplicity and accessibility, still must be found. Presumably, the best method for an advanced stream network extraction model is with the bucket-approach or a TIN approach, where the water flows over the faces of the triangles (Dakowicz & Gold, 2007; Gabrisch, 2011). The main challenge is to make these approaches also comply with the requirements of accuracy, simplicity and accessibility.

It is assumed that the reference stream that is used for verification of this research is undoubtedly the ground truth. However, it is stated that some water authorities did not yet deliver completely accurate datasets (PDOK, 2020). It is not specified what water authorities did not deliver all the data yet. To verify the results of this research more accurately, field surveying should be done. Furthermore, this research assumes, based on literature, that the D8-method with Strahler orders is the most suitable conventional and accurate approach for areas with low topographical relief. However, in future research, other conventional methods can also be executed to compare the results with the absolute point-method results. In that case, the added value of the absolute point-method can be described more sophisticatedly.

The chosen research areas and the data, self-evidently, influence the results. This research is a case study and the results could differ drastically in areas with strongly differing characteristics. As it was demonstrated that this algorithm is capable of jumping over areas with spikes in elevation, this could cause problems if there are many obstacles in the research area. By including some suggested advancements in the algorithm, this problem could be tackled. First, multiple search radii could be included in the model. Currently, only one extended search radius is included in the model, but this could be extended. Although this presumably impacts the processing speed negatively, the algorithm will be more advanced and accurate. Secondly, obstacles and other locations where the water cannot flow can be indicated with restriction lines or polygons. Also, the classification attributes can be used smartly. The usage of these classification attributes is already done in this research as only the ground points of the point cloud were used for analysis. For instance, uprights of bridges, trees or other objects can be marked as areas where the water can impossibly flow through. Lastly, the algorithm can be made more adaptive, by making the search radius dependent on the vertical profile of the area. Although these modifications presumably will not prevent the stream from flowing up as this seems impossible, given the elevation profile, it will presumably increase the accuracy of the model. Furthermore, an upward flow could be a restriction itself in this algorithm. However, this would result in premature termination of the algorithm in areas where the topographical relief is very low. The previously described trade-off between accuracy and processing time applies to this as well, as these advancements in the algorithm increase calculation times. In future research, the Python code can be made more efficient to decrease calculation times by using more efficient databases (e.g. PostGIS) and by improving the efficiency of the code. For example, currently, the lowest neighbour is identified by using an expression for field calculator as a string in the script, but this expression can also be done with Python-commands directly. This will presumably improve performance. Also, the geometries of the points are read as integers and then converted to QGIS geometries, while it is more efficient to skip this conversion step and directly use the geometries. Furthermore, it is possible to optimise the data structures of the point cloud data by using spatial indexes or by converting it to triangulations. Because of these efficient data structures and the possibility to maintain as much as possible data, the potential of stream delineation with triangulated data is high and should be further investigated (Gabrisch, 2011).

Currently, the algorithm terminates if it reaches the lowest point in the dataset or the specified outlet point. For further development, it would be better to let the calculations stop if the edge of the research area is reached. Then, there is no disturbance by low outliers in the data.

To effectively enable analysis of the LiDAR data in QGIS, this research converted the LiDAR ground-points to vector points with x, y and z values. This research shows that this is a well-functioning approach to conduct advanced analysis on point cloud data with open-source software. Even though QGIS is not created to work with such big datasets, this method adopts a smart algorithm, that only uses the necessary points for calculations. This guarantees a sufficient speed of the speed of this method. Simplicity is one of the core requirements for the created method. This includes processing time. Although the processing times in this research are satisfactory, the processing time increases significantly as the input dataset increases. Stream delineation in very big catchment areas with enormous datasets could become problematic as calculation times could become too long. However, as the point cloud data type is emerging, more and more open-source tools become available to perform direct point cloud analysis. An example is the open-source Point Data Abstraction Library (PDAL) or LAStools (Bell et al., 2021). These tools are developed to efficiently work with extremely big point cloud datasets. Currently, PDAL is a C/C++ library, which is not considered as accessible and simple in this research, but there are concrete plans to create make this library more accessible via QGIS, similarly to GDAL. By using PDALs efficient approach, the previously mentioned, more global, bucket approach (Dakowicz & Gold, 2007) or the tessellated-based surface model (Gabrisch, 2011) would become more feasible because simplicity and accessibility then improves. More above, in the next version of QGIS (v 3.18), point cloud data formats are support more extensively and it is expected that this offers a new set of openly accessible analysis tools for point clouds (Lutra Consulting, 2021). In future research, it is important to keep the scalability of this method in mind. Not only the use of efficient tools for point clouds could offer the possibility to broadly scale this method for much bigger datasets. For example, working with different levels of detail would make this method possibly scalable towards a national scale. By implementing continuous levels of detail method, points can be filtered or used adaptively to optimise computing performances (L. Zhang et al., 2020). These continuous level of detail methods are state of the art and are still in development. This research underlines the potential of broad use of point clouds and, therewith, makes the relevance for further research to efficient use of point clouds higher.

Concludingly, the results of this research show the great potential and added value of direct point cloud-analysis for different applications as this algorithm is accessible and simplistic and significantly more accurate than conventional methods. Since hydrological management applications becoming more and more important and diverse and because increasingly more LiDAR data is becoming available, more tools should be created as an addition or even a replacement for conventional tools. This can be very useful for multiple applications, also outside the field of hydrology. This LiDAR-based absolute point-method works in all types of areas, also if there is a negligible elevation difference of 0.2 meters over a horizontal distance of 700 metres. By innovatively using all the available data optimally and include smart decision rules as parameters, this algorithm offers a significantly better accuracy than conventional approaches.

Literature

Acharya, T. D., Yang, I. T., & Lee, D. H. (2018). Comparative Analysis of Digital Elevation Models between AW3D30, SRTM30 and Airborne LiDAR: A Case of Chuncheon, South Korea. 한국측량학회지, *36*(1), 17–24. https://doi.org/10.7848/KSGPC.2018.36.1.17

Aguilar, F. J., Mills, J. P., Delgado, J., Aguilar, M. A., Negreiros, J. G., & Pérez, J. L. (2010). Modelling vertical error in LiDAR-derived digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 103–110. https://doi.org/10.1016/j.isprsjprs.2009.09.003

AHN. (2019). *Kwaliteitsbeschrijving* [Web page]. AHN; AHN. /kwaliteitsbeschrijving

AHN. (2020). *Proces* [Web page]. AHN; AHN. /proces

Ai, T. (2007). The drainage network extraction from contour lines for contour line generalization. *ISPRS Journal of Photogrammetry and Remote Sensing*, *62*(2), 93–103. https://doi.org/10.1016/j.isprsjprs.2007.04.002

Anderson, D. L., & Ames, D. P. (2011). A Method for Extracting Stream Channel Flow Paths from LiDAR Point Cloud Data. *Journal of Spatial Hydrology*, 11(1).

Arun, P. V. (2013). A comparative analysis of different DEM interpolation methods. *The Egyptian Journal of Remote Sensing and Space Science*, *16*(2), 133–139. https://doi.org/10.1016/j.ejrs.2013.09.001

Aziz, N. A., Abdulrazzaq, Z., & Mansur, M. N. (2020). GIS-based watershed morphometric analysis using DEM data in Diyala River, Iraq. *The Iraqi Geological Journal*, 36–49.

Bell, A., Chambers, B., & Butler, H. (2021, januari 22). *PDAL* [Documentation]. https://pdal.io/about.html

Chen, Q. (2007). Airborne Lidar Data Processing and Information Extraction. *Photogrammetric Engineering*, *73*(2), 109–112.

Clevers, J. (2020). *Remote Sensing Theory Syllabus*. http://www.geo-informatie.nl/courses/gima_rs/

CloudCompare. (2020). *Normals\Compute*. https://www.cloudcompare.org/doc/wiki/index.php?title=Normals%5CCompute#Description

Dakowicz, M., & Gold, C. (2007). Finite difference method runoff modelling using Voronoi cells. *Proceedings, 5th. ISPRS Workshop on Dynamic and Multi-Dimensional GIS (DMGIS)*, 55–60.

de Azeredo Freitas, H. R., da Costa Freitas, C., Rosim, S., & de Freitas Oliveira, J. R. (2016). Drainage networks and watersheds delineation derived from TIN-based digital elevation models. *Computers & Geosciences*, *92*, 21–37.

Dyn, N., Levin, D., & Rippa, S. (1990). Data dependent triangulations for piecewise linear interpolation. *IMA Journal of Numerical Analysis*, *10*(1), 137–154.

Esri. (2020). *How Kriging works—Help | ArcGIS for Desktop* [Informational page]. https://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.htm

Esri. (2021). *How Surface Contour works—ArcGIS Pro | Documentation* [Informational page]. https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-surface-contour-works.htm

Gabrisch, G. B. (2011). *Irregular tessellated surface model map algebras to define flow directions and delineate catchments using LiDAR bare earth sample points*.

Garbrecht, J., & Martz, L. W. (2000). Digital Elevation Model Issues in water resources Modeling. *Hydrologic and hydraulic modeling support with geographic information systems*, 1–28.

GIM International. (2020). Photogrammetry and Point Clouds. https://www.gim-international.com/themes/photogrammetry-point-clouds

Gold, C. (2016). Tessellations in GIS: Part I—putting it all together. *Geo-Spatial Information Science*, *19*(1), 9–25. https://doi.org/10.1080/10095020.2016.1146440

Gortzak, I., van Rosmalen, R., Ticheloven, S., & Verwer, J. (2020). *Identifying the added value of point cloud representation in stream delineation modelling in low elevation areas*. Utrecht University.

Hjelle, Ø., & Dæhlen, M. (2006). *Triangulations and applications*. Springer-Verlag.

Höfle, B., & Rutzinger, M. (2011). Topographic airborne LiDAR in geomorphology: A technological perspective. *Zeitschrift Für Geomorphologie, Supplementary Issues*, *55*(2), 1–29. https://doi.org/10.1127/0372-8854/2011/0055S2-0043

Hui, Z., Hu, Y., Yevenyo, Y., & Yu, X. (2016). An Improved Morphological Algorithm for Filtering Airborne LiDAR Point Cloud Based on Multi-Level Kriging Interpolation. *Remote Sensing*, *8*(1), 35. https://doi.org/10.3390/rs8010035

IHE Delft Institute for Water Education. (2020). *Research lines*. https://www.un-ihe.org/chair-groups/hydrology-water-resources/research-lines

Johnson, K., Nissen, E., Saripalli, S., Arrowsmith, J. R., McGarey, P., Scharer, K., Williams, P., & Blisniuk, K. (2014). Rapid mapping of ultrafine fault zone topography with structure from motion. *Geosphere*, *10*(5), 969–986. https://doi.org/10.1130/GES01017.1

Kolingerová, I. (2004). On Triangulations. In A. Laganá, M. L. Gavrilova, V. Kumar, Y. Mun, C. J. K. Tan, & O. Gervasi (Red.), *Computational Science and Its Applications – ICCSA 2004* (Vol. 3044, pp. 544–553). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-24709-8_58

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic information science and systems*. John Wiley & Sons.

Lutra Consulting. (2021, februari 18). *Support for Point Clouds Released in QGIS 3.18—Lutra Consulting* [Informational page]. https://www.lutraconsulting.co.uk/blog/2021/02/18/qgis-3-18-point-cloud/

Marsh, C. B., Spiteri, R. J., Pomeroy, J. W., & Wheater, H. S. (2018). Multi-objective unstructured triangular mesh generation for use in hydrological and land surface models. *Computers & Geosciences*, *119*, 49–67. https://doi.org/10.1016/j.cageo.2018.06.009

Meesuk, V. (2017). *Point Cloud Data Fusion for Enhancing 2D Urban Flood Modelling*. CRC Press.

Nelson, E. J., Jones, N. L., & Miller, A. W. (1994). Algorithm for Precise Drainage-Basin Delineation. *Journal of Hydraulic Engineering*, *120*(3), 298–312. https://doi.org/10.1061/(ASCE)0733-9429(1994)120:3(298)

PDOK. (2020). *Metadata Hydrography*. http://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/07575774-57a1-4419-bab4-6c88fdeb02b2?tab=general

PDOK. (2021, januari 29). *Waterschappen Oppervlaktewateren IMWA*. https://service.pdok.nl/hwh/oppwaterimwa/atom/v1_0/oppwaterimwa.xml

Pfeifer, N. (2002). *3D terrain models on the basis of a triangulation*. Inst. für Photogrammetrie u. Fernerkundung d. Techn. Univ.

QGIS API Documentation. (2020). [Handbook]. QGS TriangularMesh. https://qgis.org/api/qgstriangularmesh_8h_source.html#l00049

QGIS Documentation. (2020). [Handbook]. Writing new Processing algorithms as Python scripts —. https://docs.qgis.org/3.4/en/docs/user_manual/processing/scripts.html

Qu, G., Su, D., & Lou, Z. (2014). A new algorithm to automatically extract the drainage networks and catchments based on triangulation irregular network digital elevation model. *Journal of Shanghai Jiaotong University (Science)*, *19*(3), 367–377. https://doi.org/10.1007/s12204-014-1511-9

Rheinwalt, A., Goswami, B., & Bookhagen, B. (2019). A Network-Based Flow Accumulation Algorithm for Point Clouds: Facet-Flow Networks (FFNs). *Journal of Geophysical Research: Earth Surface*, *124*(7), 2013–2033. https://doi.org/10.1029/2018JF004827

Rodríguez, N., & Silveira, R. (2017). Implementing Data-Dependent Triangulations with Higher Order Delaunay Triangulations. *ISPRS International Journal of Geo-Information*, *6*(12), 390. https://doi.org/10.3390/ijgi6120390

Si, H. (2020). *TetGen—A Quality Tetrahedral Mesh Generator and 3D Delaunay Triangulator Version 1.6. 1.6*.

Szczepaniak-Kołtun, Z. (2015). Use of lidar data and selected algorithms for determining the flow lines. *Technical Sciences*, 11.

Ujaval Gandhi. (2020, november 27). *QGIS with Ujaval Gandhi*. https://www.youtube.com/watch?v=IraiAF4s5Tg&feature=youtu.be&t=2194

van Bemmelen, J., Quak, W., van Hekken, M., van Oosterom, P., & Box, P. O. (1993). *Vector vs. Raster-based Algorithms for Cross Country Movement Planning*. 14.

van der Kwast, H., & Menke, K. (2019). *QGIS for Hydrological Applications*.

Verbree, E., & van Oosterom, P. (2001). Scanline forced Delaunay TENs for surface representation. *International Archives of Photogrammetry and Remote Sensing, Volume XXXIV-3/W4 Annapolis*.

Verbree, E., & van Oosterom, P. (2003). Better surface representations by delaunay tetrahedronized irregular networks. *ESRI International User Conference*, 24.

Wang, L., & Liu, H. (2006). An efficient method for identifying and filling surface depressions in digital elevation models for. *International Journal of Geographical Information Science*, *20.2*, 22.

Wang, Z., Ma, H., Zheng, C., & Chen, R. (2018). *High-Quality Contour Line Generation from LiDAR Point Clouds for the Area of Forests* [Preprint]. MATHEMATICS & COMPUTER SCIENCE. https://doi.org/10.20944/preprints201802.0003.v1

Xiaoye Liu. (2008). Airborne LiDAR for DEM generation: Some critical issues. *Progress in Physical Geography: Earth and Environment*, *32*(1), 31–49. https://doi.org/10.1177/0309133308089496

Yang, B., Dong, Z., Liu, Y., Liang, F., & Wang, Y. (2017). Computing multiple aggregation levels and contextual features for road facilities recognition using mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *126*, 180–194. https://doi.org/10.1016/j.isprsjprs.2017.02.014

Zhang, F., Zhou, Q., Li, Q., Wu, G., & Liu, J. (2018). An enhanced approach for surface flow routing over drainage-constrained triangulated irregular networks. *Transactions in GIS*, *22*(1), 43–57. https://doi.org/10.1111/tgis.12294

Zhang, L., van Oosterom, P., & Liu, H. (2020). Visualization of point cloud models in mobile augmented reality using continuous level of detail method. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIV-4/W1-2020*, 167–170. https://doi.org/10.5194/isprs-archives-XLIV-4-W1-2020-167-2020

Zhou, Q., & Chen, Y. (2011). Generalization of DEM for terrain analysis using a compound method. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(1), 38–45. https://doi.org/10.1016/j.isprsjprs.2010.08.005

Zhu, Q., Feng, J., & Huang, J. (2016). Natural neighbor: A self-adaptive neighborhood method without parameter K. *Pattern Recognition Letters, 80*, 30–36. https://doi.org/10.1016/j.patrec.2016.05.007

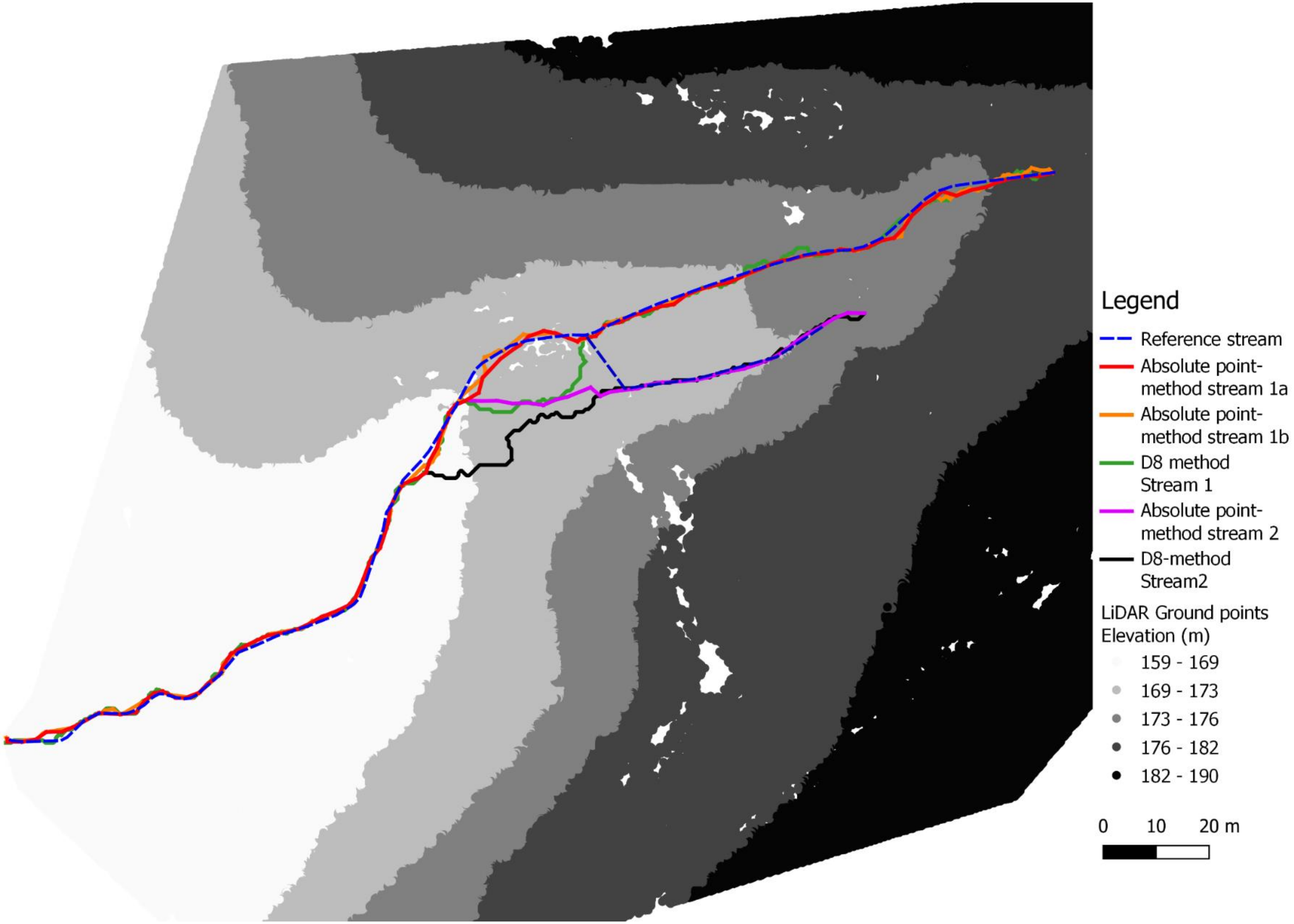
Appendix A: Stream Delineation with QGIS Python script

Please consider the attached file: ‘stream_delineation.py’ or via: https://github.com/stn228/Absolute_point_method_Stream_delineation

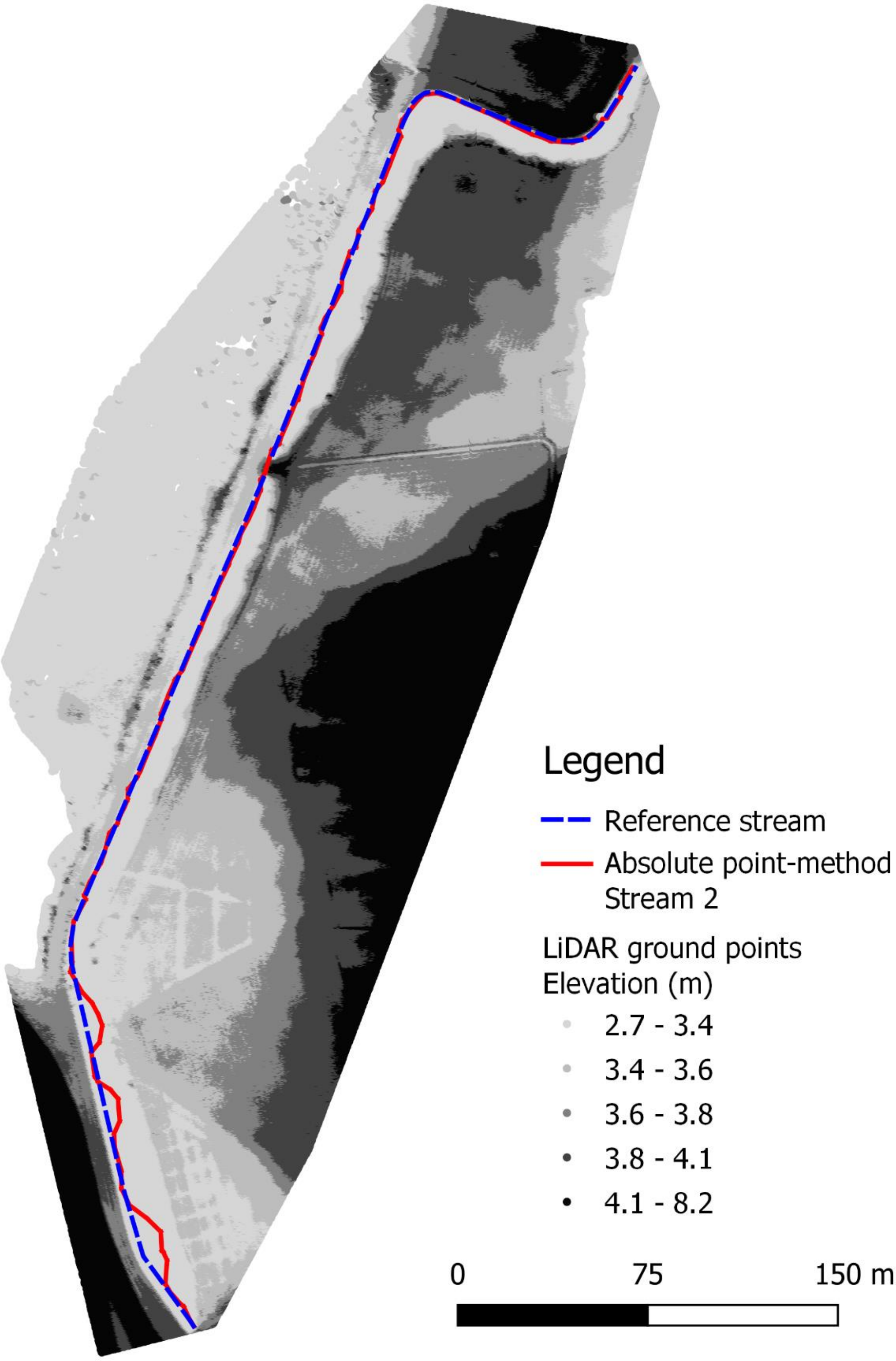
Appendix B: Python script to plot the longitudinal profiles

Please consider the attached file: ‘create_graphs_Epen.py’ and ‘create_graphs_Olst.py’

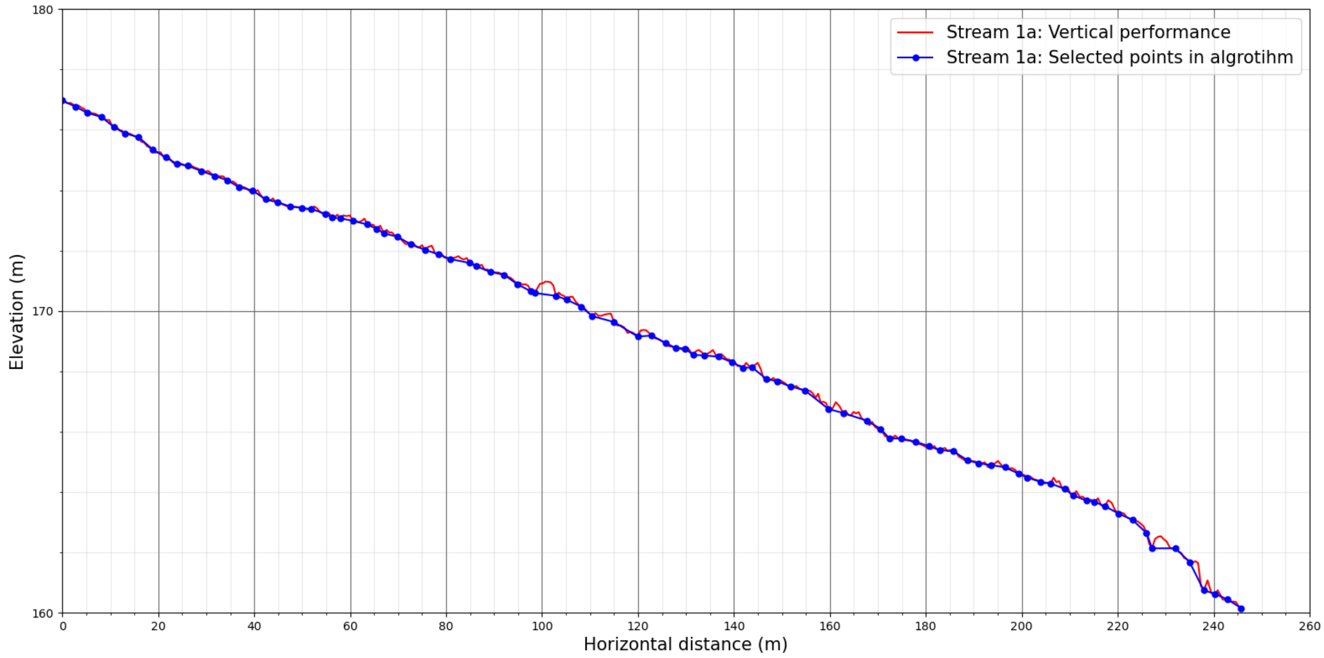
Appendix C: Overview of the results in the Epen research area



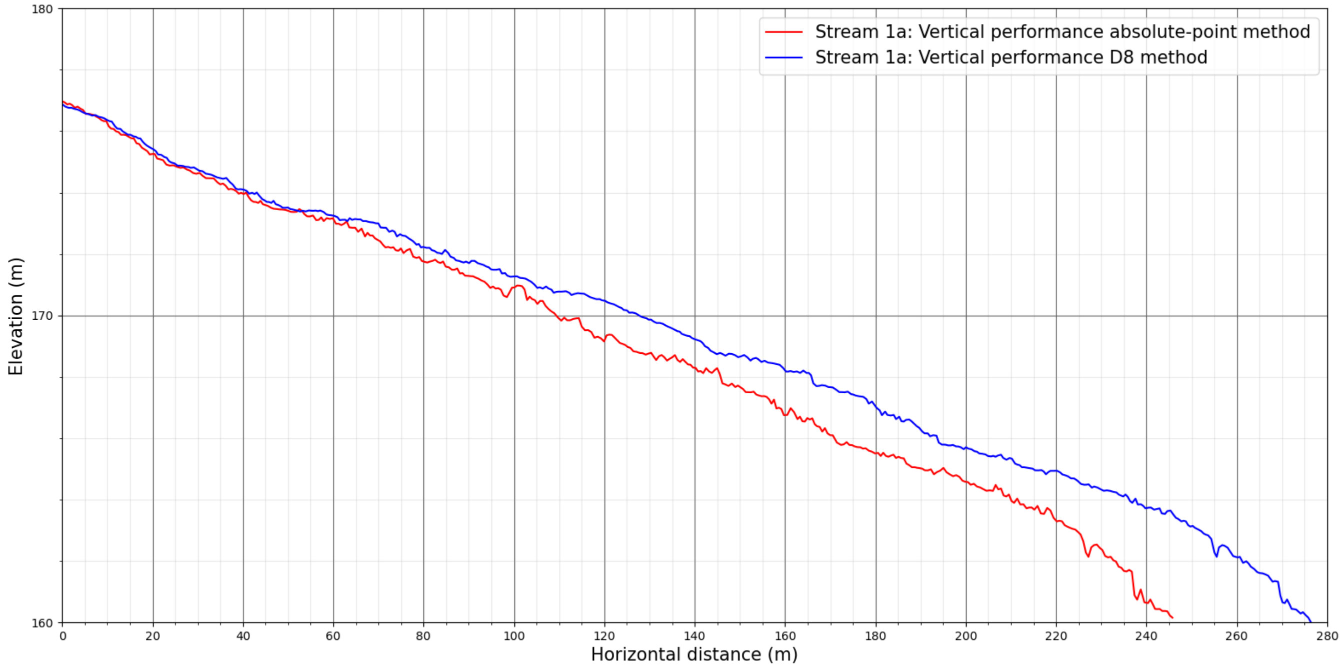
Appendix D: Overview of the results in the Olst research area



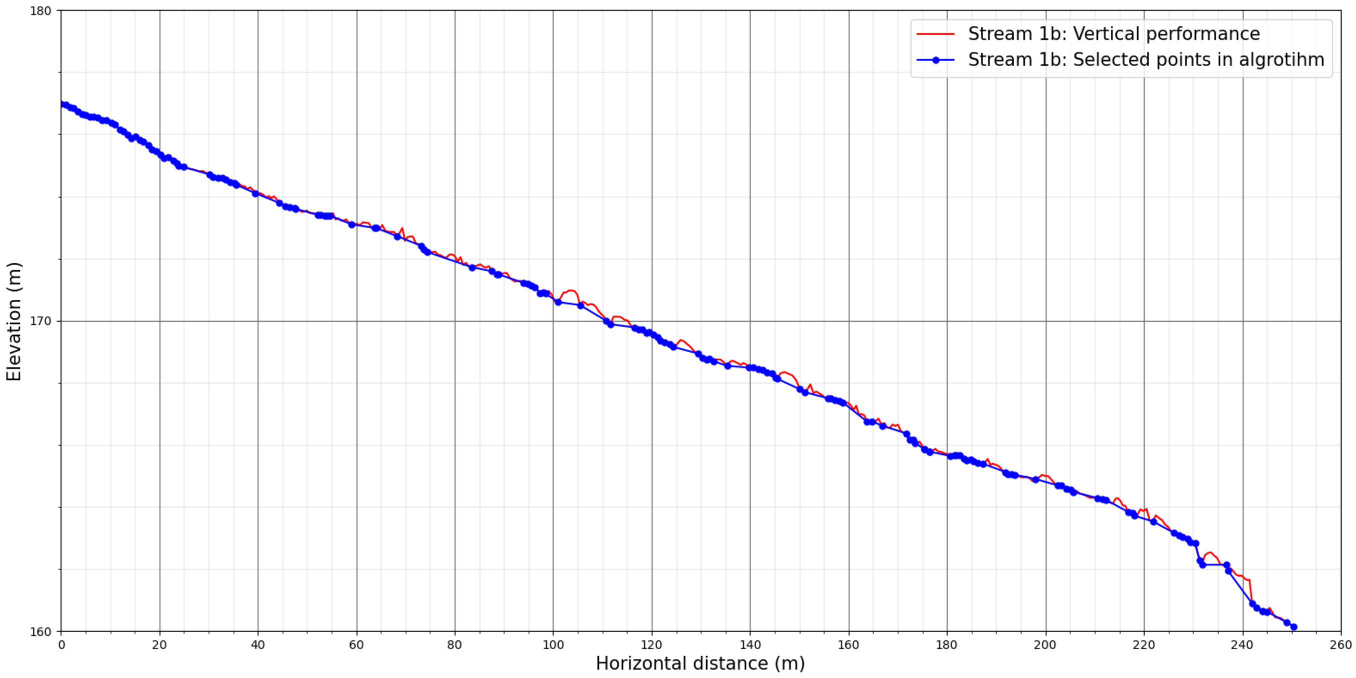
Appendix E: Overview of the vertical performance of all the modelled streams



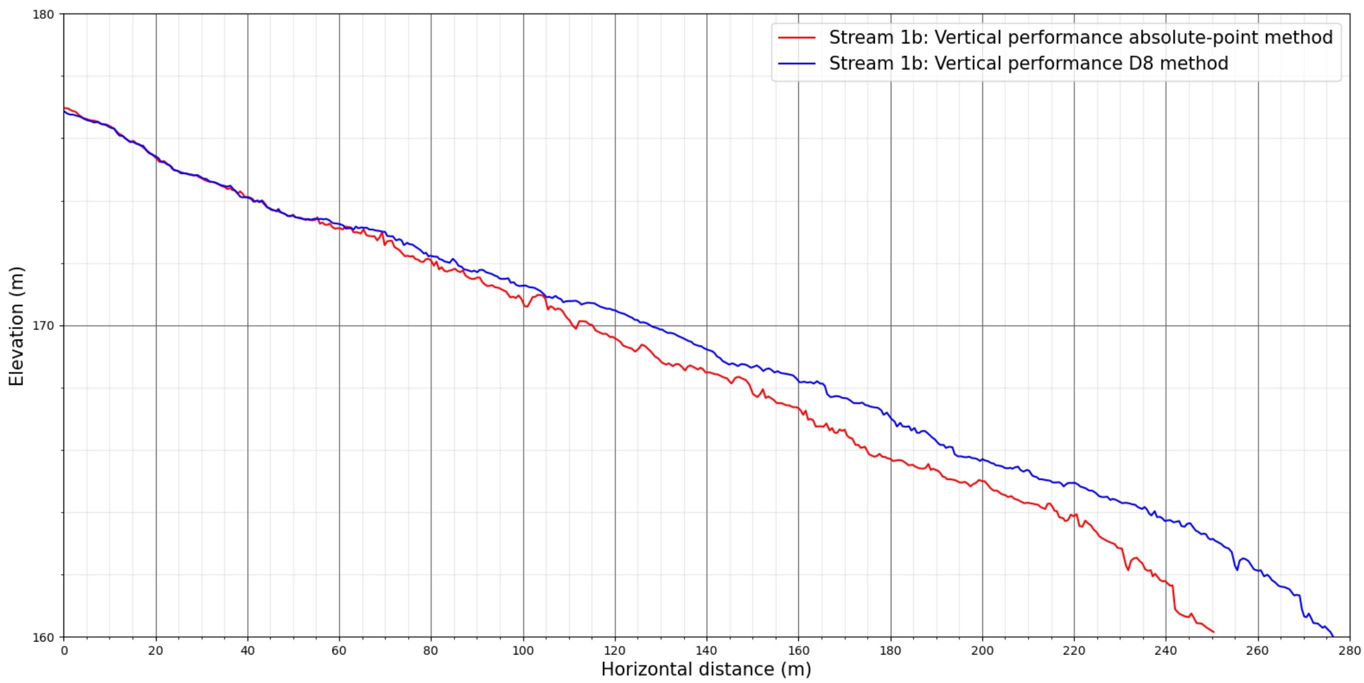
Stream 1a: vertical performance compared to the selected points by the absolute point-method



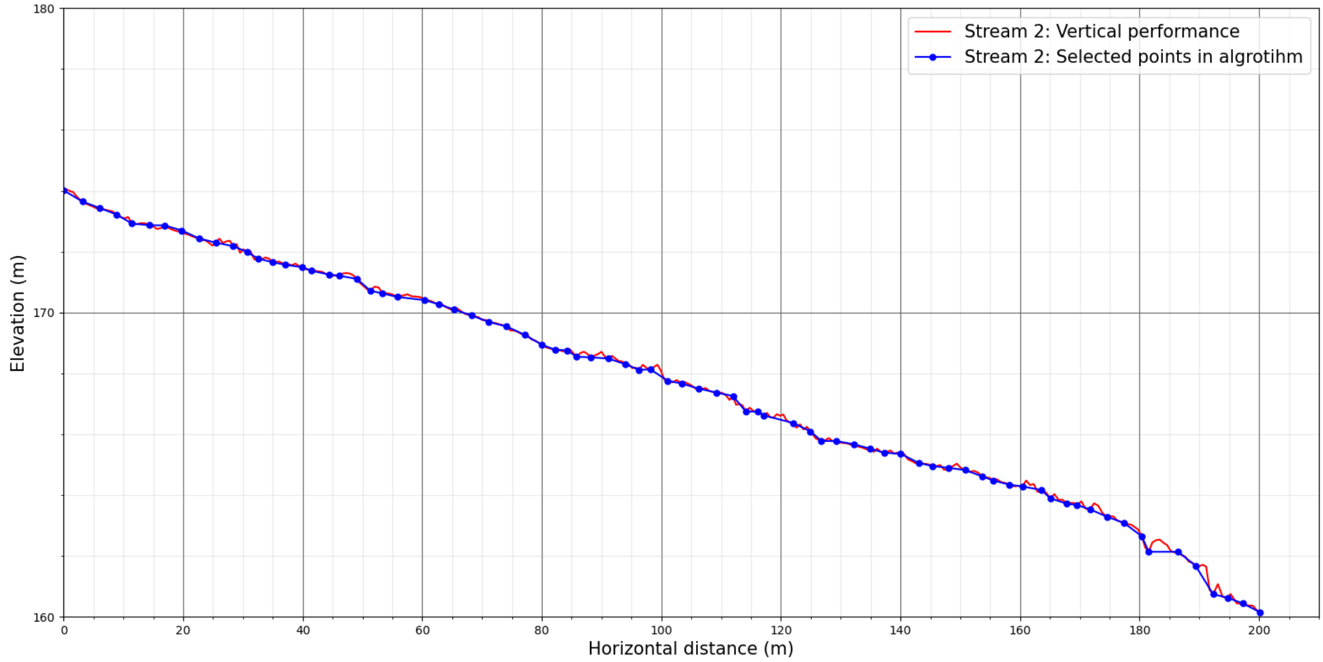
Stream 1a: The results of the D8-method and the absolute point-method compared.



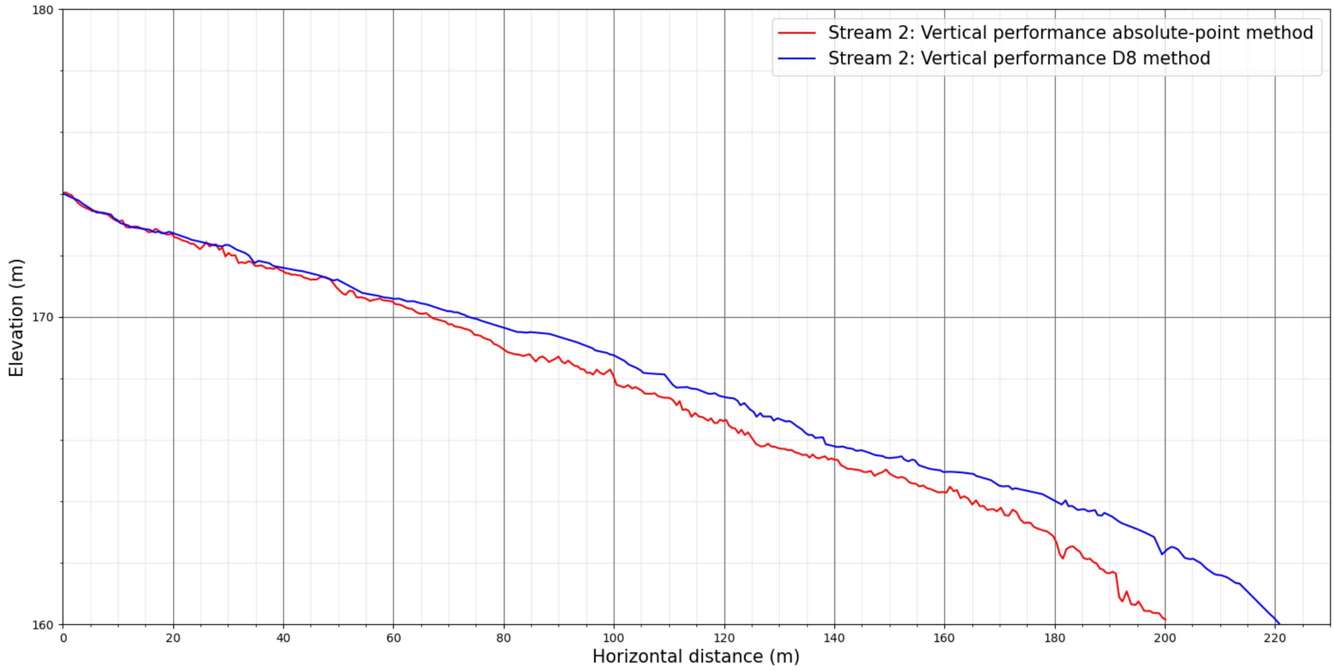
Stream 1b: vertical performance compared to the selected points by the absolute point-method



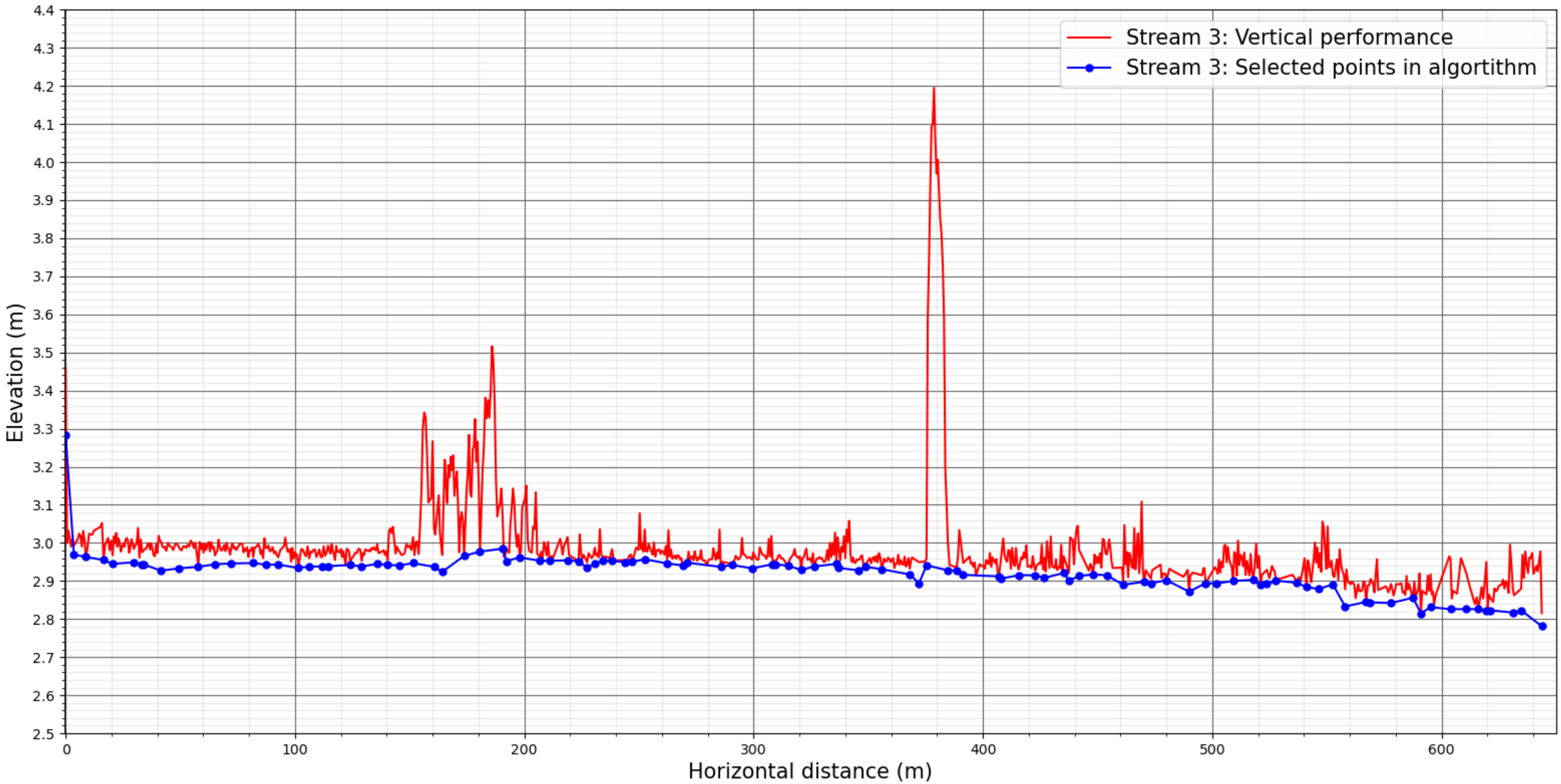
Stream 1b: The results of the D8-method and the absolute point-method compared.



Stream 2: vertical performance compared to the selected points by the absolute point-method



Stream 2: The results of the D8-method and the absolute point-method compared.



Stream 3: vertical performance compared to the selected points by the absolute point-method