Improvement of automatic generalisation of manmade water networks for topographic maps by context-dependent pruning

Dissertation by Vincent van Altena

Submitted in part fulfilment of the requirements for the degree of Master of Science in Geographical Information Systems (UNIGIS), Faculty of Economics and Business Administration, VU University Amsterdam, The Netherlands.

December 2014

VU AMSTERDAM
GEOGRAPHICAL INFORMATION SCIENCES

UNIGIS
FEWEB-RE
Vrije Universiteit
De Boelelaan 1105
1081 HV Amsterdam
+31 (0)20-5986099
unigis@feweb.vu.nl
www.unigis.nl

TITLE	Improvement of automatic generalisation of manmade water networks for topographic maps by context-dependent pruning
	Dissertation submitted in part fulfilment of the requirements for the degree of Master of Science in Geographical Information Systems (UNIGIS), Faculty of Economics and Business Administration, VU University Amsterdam, The Netherlands
DATE	December 2014
NAME	Vincent van Altena
STUDENT NUMBER	1888323
FIRST SUPERVISOR	prof. dr. J.E. Stoter (Technische Universiteit Delft)
SECOND SUPERVISOR	prof. dr. H.J. Scholten (Vrije Universiteit Amsterdam, UNIGIS)

~

This page was intentionally left blank

~

Abstract

The issue of automated generalisation resounds for decades in the cartographic and academic world and has been envisaged as the 'holy grail' of cartography (Anderson-Tarver et al. 2011). Recently, there have been several achievements, such as the introduction of OS MasterMap in Great Britain and the replacement of the manual generalisation production line by a fully automated workflow at the Dutch Kadaster. Despite these successes, it is also acknowledged that further development is necessary for certain subjects and situations. One of these areas is the pruning of manmade water networks. Prevailing water thinning algorithms do not deliver satisfying results, because they do not handle the cultivated manmade water network correctly.

The focus of this study is to research existing, and to develop new alternative thinning methods which account for landscape types. Based on effectiveness for each separate landscape type, advice will be given for improvement of water thinning in the Dutch situation. On a more abstract level the study will deliver a methodology for pruning of artificial manmade networks with regard to landscape typology and will research methods for evaluation of the quality of generalisation results.

A better understanding of thinning, landscape and evaluation is gained from literature. It appears several thinning algorithm exist, each with its own distinguishing focus. Literature also provides an evaluation framework, but not all suggested methods are applicable in this research.

The concepts from literature are tested and evaluated by several experiments. The first experiment start with the identification of landscape variation for the dataset. It is possible to distinguish landscape type for areas based on feature morphology and humidity, as several experiments show. Test areas are selected which represent one of the identified landscape types. In a subsequent experiment, the concept of geometric network improvement is tested and results show significant connectivity increase for each test area.

The final experiments researched the effectivity on differing landscape types of the three alternative thinning algorithms. The results of the algorithms are judged based on the amount of thinning, the resemblance of the results with the input data and on the deviation in connectivity. These results are judged in coherence, for both the individual as well as the total of the test areas. It appears that the Thin Road Network-algorithm provides best results for all landscape types. However, there is some variation distinguishable for the different test areas.

The results of this research can be used in further investigations for tailoring the method of thinning to the identified landscape type. The proposed metrics to measure the effectivity of thinning algorithms, reduction, resemblance and connectivity, provide a good basis for comparison of results of alternative approaches.

Samenvatting

In de cartografische en academische wereld speelt het onderwerp automatische generalisatie al enkele decennia en het is wel aangemerkt als 'de heilige graal' van de cartografie (Anderson-Tarver et al. 2011). Recentelijk hebben zich verscheidene ontwikkelingen voorgedaan, zoals de introductie van OS MasterMap in Groot Brittannië en de vervanging van de handmatige generalisatie-productielijn door een volledig automatisch systeem bij het Nederlandse Kadaster. Ondanks deze successen wordt ook erkend dat er nog meer ontwikkeling nodig is voor bepaalde onderwerpen en situaties. Eén van deze onderwerpen is de uitdunning van kunstmatige waternetwerken. Bestaande water-uitdunningsalgoritmen leveren geen bevredigende resultaten, omdat ze een kunstmatig, aangelegd waternetwerk niet correct kunnen verwerken.

De focus van deze studie is het onderzoeken van bestaande, en het ontwikkelen van nieuwe alternatieve uitdunningsmethoden die rekening houden met landschapstypering. Op basis van de effectiviteit voor elk landschapstype afzonderlijk zal er advies worden gegeven hoe de wateruitdunning voor de Nederlandse situatie verbeterd kan worden. Op een meer abstract niveau zal het onderzoek een methodologie voor de uitdunning van kunstmatig aangelegde netwerken opleveren die rekening kan houden met landschapstypering; en zullen methoden voor de evaluatie van de kwaliteit van generalisatie resultaten worden onderzocht.

Een beter begrip van uitdunning, landschap en evaluatie wordt door literatuurstudie verkregen. Het blijkt dat er verschillende uitdunningsalgoritmen bestaan, die elk een eigen focus hebben. De literatuur voorziet daarnaast in een evaluatie-raamwerk, maar helaas zijn niet alle gesuggereerde methoden bruikbaar binnen dit onderzoek.

De concepten van de literatuurstudie zijn door middel van enkele experimenten getest en geëvalueerd. Het eerste experiment identificeert landschapsvariatie in de dataset. Enkele experimenten tonen aan dat het mogelijk is om het landschapstype van gebieden te baseren op morfologie en vochtigheidsgraad (*humidity*). Het onderwerp van een volgende test is geometrische verbetering van een netwerk. De resultaten daarvan tonen een significante verbetering in connectiviteit voor elk testgebied.

De laatste groep experimenten heeft de effectiviteit van de drie uitdunningsalgoritmen voor de verschillende landschapstypen onderzocht. De resultaten van de algoritmen zijn beoordeeld op de mate van uitdunning, de gelijkenis van de resultaten met de inputdata en op de wijziging in connectiviteit. De resultaten van deze drie criteria zijn in samenhang beoordeeld, zowel voor de testgebieden afzonderlijk als voor het geheel aan testgebieden. Hieruit blijkt dat het Thin Road Network-algoritme de beste resultaten geeft voor alle landschapstypen. Er zit echter wel variatie in de resultaten voor de verschillende testgebieden.

In een vervolgstudie kunnen de resultaten van dit onderzoek gebruikt worden om de keus voor het toe te passen uitdunningsalgoritme te baseren op berekende landschapscriteria. De voorgestelde criteria om de effectiviteit van uitdunningsalgoritmen te meten, de mate van uitdunning, de gelijkenis met het origineel en het behoud van connectiviteit, geven wanneer ze in samenhang worden beschouwd inzicht om resultaten van alternatieven te kunnen vergelijken.

~

This page was intentionally left blank

~

Contents

List of tables

List of illustrations

Disclaimer

The results presented in this thesis are based on my own research at the Faculty of Economics and Business Administration of the VU University Amsterdam. All assistance received from other individuals and organisations has been acknowledged and full reference is made to all published and unpublished sources. This thesis has not been submitted previously for a degree at any institution.

Signed:

Zwolle, 01-12-2014

V.P. van Altena BTh

Preface

This project was conceived during my time working in the project on automatic generalisation for Kadaster. As a GIS specialist, I was closely involved in the practical research and development of a production line for automatic generalisation of topographic maps from 10:000 to 50:000 scale. During this research a lot of encountered issues were tackled, while others are still in need of solution. One of these issues is the pruning of water.

My main reason for choosing this topic is personal dissatisfaction with current results, but also the drive to research some ideas (from literature and in prototyping) which could improve the outcomes. As time schedules from every day work are often overloaded, doing a master thesis is an excellent opportunity to spend a fair amount of time on this research. To quote Horace (Ars Poetica, 343 (8AD)): "*Miscemus utile dulci*"[1].

Acknowledgements

I would like to thank my employer, Kadaster for having me as a member of the research team for Automatic Generalisation and facilitating me to do a Master study parallel to work. A particular *thank you* goes to Ben Bruns, my direct manager, for having the confidence I would complete my thesis and for all his *stimuli* to continue my studies. My close colleagues, Ron Nijhuis, Marc Post, Jan Bakermans and Hedwig Kupers; I owe you a lot. Working together on the automatic generalisation project at Kadaster meant a lot of fun and frustration, but together we endured. Without our thought-provoking co-operation, the project would not have reached its current state.

Thank you to the Vrije Universiteit (VU) in Amsterdam for allowing me the freedom to have prof. dr. Jantien Stoter from Delft Technical University as first supervisor. Jantien, thank you for all guidance and critiques during my research project and helping me to keep focus in delivering an academic study. I also acknowledge prof. dr. Henk Scholten (Program Board Chair) for his support and practical approach of writing thesis proposals and dr. Niels van Manen (MSc Programme Co-ordinator). In addition I would like to thank all VU – UNIGIS staff for logistical and administrative support.

Alexander Albersz and Paul Hardy also deserve to be mentioned here. You both contributed a lot to the end result by proof reading and correcting my English. Thank you both, I owe you one!

A special word of thanks goes to my family. Dear Roy and Jesse, both of you made an immeasurable contribution to my studies in just being who you are. My beloved wife, Ruth, your support, endurance with me and just loving me for who I am, have been a driving force to finish my studies (Prov. 12:4a & 18:22)!

[1] "to mingle the useful with the pleasant"

1 Introduction

This thesis studies the automatic pruning of artificial hydrographical networks. A specific problem which occurs in the automated deduction of overview maps from detailed maps, called "generalisation". The main terminology is explained in the following paragraphs.

Map making is known as a common pursuit of mankind. Ancient Babylonians and Chinese were already producing maps, centuries before the start of our era. Just like nowadays, they made choices in how to convey the important characteristics of an area. Such decisions are always made at the expenses of inferior characteristics to provide legibility of the end result. Whether the result is successful, depends on the success in delivering a message from designer to user. For topographic maps the purpose is to convey a faithful representation of a landscape for various applications such as navigation or logistical planning. The involved discipline is called *Cartography*.

1.1 Cartography

In the quest for a definition of cartography, Kraak (2010) notes that the usage of the term changed significantly. The term *cartography* was first used simply to indicate the *production of maps*. However, since the 1960's, the emergence of digital techniques within the field of cartography influenced its understanding and definition.

This development is recognisable in the definitions of Kraak (2010), Guptill & Starr (1984) and Taylor (1991). All three stress *usage*, *digital techniques* (i.e. storage in databases) and *variance in media* (i.e. paper, digital devices) as typical characteristics of modern cartography. The International Cartographic Association (ICA) defines cartography as 'the discipline dealing with the art, science and technology of making and using maps'. (Schmidt 2014)

While various definitions of the term Cartography have been suggested, this thesis will use the definition suggested by the ICA. It should however be noted that when using terms as *cartography*, *cartographer*, *cartographical* (*-ly*) the focus in this thesis is on the *making* of maps and not on their utilisation.

1.2 Map

What is a map? There have been several attempts to define the term. Kraak (2010, 41) offers a few definitions:

- a map is 'a graphic model of the geospatial aspects of reality';
- a map is 'a conventional image, mostly on a plane, of concrete or abstract phenomena which can be located in space';
- 'a representation or abstraction of geographic reality. A tool for presenting geographic information in a way that is visual, digital or tactile'.

He also cites the definition of the ICA which defines map as

- 'a symbolised representation of a geographical reality, representing selected features and characteristics, resulting from the creative effort of its author's execution of choices, and designed for use when spatial relationships are of primary relevance'. (Schmidt 2014)

All four definitions of ‘map’ stress the modelling of certain geographical aspects of reality which is studied in the discipline of Topography.

Topography is a contraction and its etymology leads back to the Greek words *topos* (place) and *graphia* (description). The discipline of topography is concerned with the “arrangement of the natural and artificial physical features of an area” (Oxford dictionary 2014). Consequently, topographic maps give “a detailed description or representation on a map of the physical features of an area”(Oxford dictionary 2014).

The creation of topographic maps inevitably involves visual selection, symbolisation and abstraction to provide information for specific use cases. These use cases not only dictate the contents of a map, but also the extent of the geographical area it is representing. Almost all maps portray objects which are a reduction in size of the real geographical objects, to be able to provide large areas on a single map sheet. This brings in the conception of *scale* which is the “ratio between a distance on a map and the corresponding distance in the terrain.“ (Kraak 2010, p.41). This is an important concept with regard to generalisation.

1.3 Generalisation

Reduction of scale should not be narrowed to reduction in zoom factor, as is illustrated by Mackaness, Ruas, and Sarjakoski (2007): ‘Geographers have long understood the relevance of scale to the discernment of pattern and the derivation of meaning from geographic data. The power of the map lies in its ability to abstract space. It plays a critical role in identifying and interpreting process.’(2007, p.1)

The abstraction of space in making maps includes exaggeration of important, and exclusion of marginal elements, which is illustrated in Figure 1 and Figure 2.

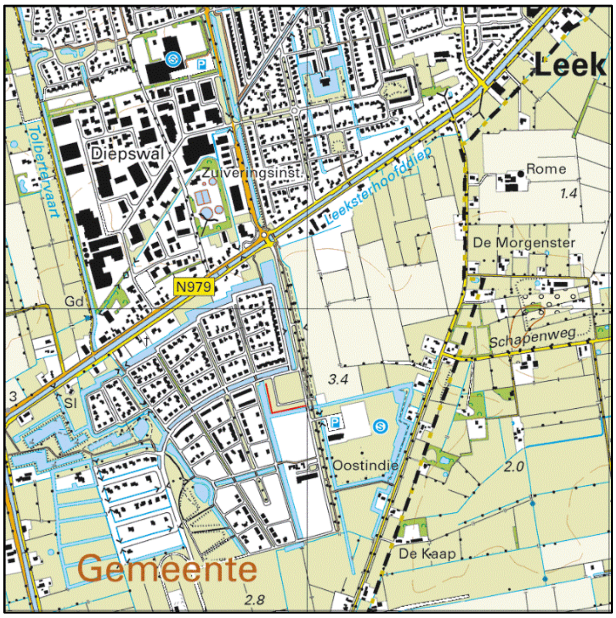


Figure 1: a fragment of a 25k map

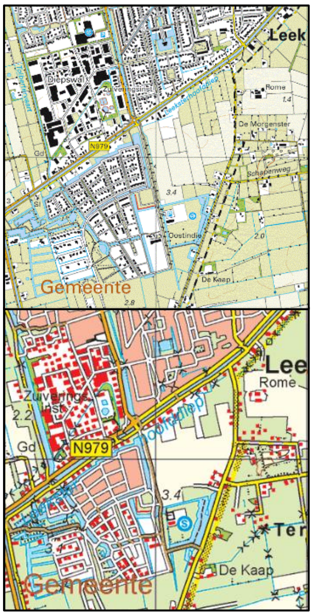


Figure 2: above the same 25k map fragment reduced in map scale; below the same geographical area, but portrayed as a generalised 50k map fragment

Both figures indicate the same issues. A simple reduction in scale is not sufficient: an overload of details makes the reduced map illegible. An intelligent discernment of geographical patterns is frustrated by a lack of emphasis. To portray this issue in Kraaks terms: “Each map within a certain *scale* range requires its own level of detail depending on its purpose.[…] The process of reducing the amount of detail in a map in a meaningful way is called generalisation. The process of generalisation is normally executed when the map scale has to be reduced.**”**(2010, p.95)

The International Cartographic Association (ICA) defines generalisation as: “…the selection and simplified representation of detail appropriate to the scale and /or the purpose of a map…” (International Cartographic Association - Commission II 1973, p.173)

Kraak also points out that ‘generalisation entails information loss, but one should try to preserve the essence of the contents of the original map. This implies maintaining geometric and attribute accuracy, as well as the aesthetic quality of the map’ (2010, p.97).

1.3.1 Kinds of generalisation

Grünreich (1988) was the first to differentiate model and map generalisation with regard to Digital Landscape Models and Digital Cartographic Models: Kraak (2010, p.96) expanded this framework and distinguishes three kinds of generalisation. Both views are supplementary and are summarized as follows:

- When reality is translated into a Digital Landscape Model, selection and classification occurs (specified objectively in formal regulations or subjectively in the mind of the data acquirer). This process of abstracting objects from reality is called *object generalisation*’.
- *Model generalisation* relates to the translation from one DLM to another DLM. The application of more abstraction removes elements from (multiple) layers. The generalisation operators used here are elimination and reclassification.
- *Cartographic generalisation* simplifies resulting images to remove graphic conflicts. Cartographic generalisation is traditionally divided into two subclasses: graphic versus conceptual generalisation. The latter is primarily applicable on thematic maps (i.e. geological maps) and requires knowledge of the themes to be generalised, while the former is more focussed on topographic maps.

The framework can be seen as a system of direct and indirect lenses on reality, as is shown in the figure below:

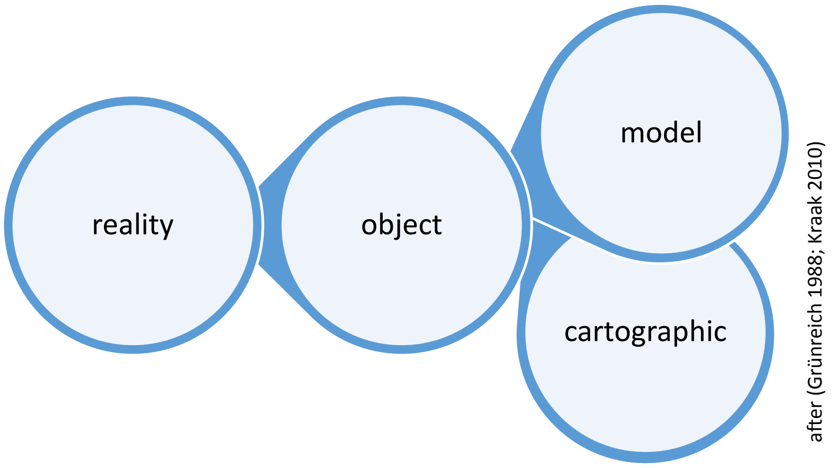


Figure 3: framework for abstraction levels in generalisation

1.4 Automatic Generalisation

Generalisation has been the sphere of work of many craftsmen until recently. *Automatic* generalisation in production environments has a relatively short history. In 1997 Peng still was forced to conclude that …

> "After more than three decades of effort, it is still a question whether generalisation can be formally defined, and whether automated generalisation can be realized." (1997, p.i),

… but there have been developments which have drastically changed the domain. Over the years several academic researches have been published in the field of automatic generalisation. (Burghardt et al. (2014); Stoter (2010a); Stoter (2010b); Mackaness et al. (2007) can be consulted for an overview of the state-of-the-art until 2014). That generalisation research has matured is shown by the implementation of generalisation solutions in commercial software such as Esri and 1Spatial and (partial) implementations at National Mapping Agencies, thus replacing traditional manual labour and transforming existing production lines (Duchêne et al. 2014; ICA Commission on Generalisation and Multiple Representation 2013). Several National Mapping Agencies (NMAs) have introduced automatic generalisation in their production workflows to partially replace parts of manual labour: see e.g. Turkey (Bildirici 2004), Catalonia (Baella & Pla 2005), China (Yan et al. 2006), France (Lecordix et al. 2007), Denmark (West-Nielsen & Meyer 2007), Great Britain(Revell et al. 2011), Italy (Garnero & Godone 2012), USA, Germany, and Switzerland. (Duchêne et al. 2014)

Beside this progress in research and implementation, also external factors can be identified. Due to the economic crisis, several NMAs were facing budget cuts. They started rethinking their production methods and quality standards. In some cases this caused a paradigm change about quality: no longer cartographic and aesthetic quality are the ultimate goal, but fitness for users now has become the leading objective, as is illustrated by the adage "good is good enough". The increased use of digital geographical information (i.e. on the internet or mobile devices) and the associated expectations on up-to-date-ness can also be regarded as giving impetus to this paradigm change.

1.4.1 Application of Automated Generalisation by National Mapping Agencies

Duchêne et al. (2014) propose at least two classifications to distinguish the NMAs' approaches to automatic generalisation: (1) the *period of duration* to encourage research and a subsequent transfer to production; and (2) the *actor in* and *approach of* the actual implementation into a production line, as is shown in Figure 4.

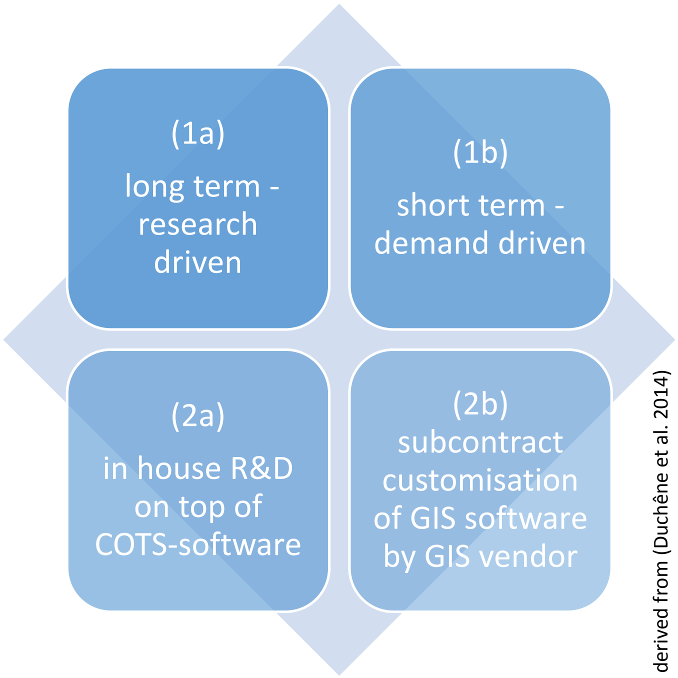


Figure 4: classification matrix of NMA efforts with regard to automatic generalisation

The addition of the thoroughness of implementation (partial vs. full) adds another dimension to this classification matrix, which can be used to classify the current state of automatic generalisation within NMAs, see Table 1.

Table 1: classification of the current state of automatic generalisation within NMAs

NMA	PERIOD		ACTOR & APPROACH		THOROUGHNESS		REFERENCE
	long	short	in house	sub-cont.	partial	full	
ICC (CATALUNYA)							(Baella & Pla 2014)
IGN (FRANCE)					*	**	(Lecordix & Maugeais 2014)
ORDNANCE SURVEY (UK)						**	(Revell et al. 2011; Regnauld 2014)
KADASTER (NL)							(Stoter, Post, et al. 2014)
SWISSTOPO (SCHWITZERLAND)							(Käuferle 2014)
USGS (USA)							(Stanislawski, Brewer, et al. 2014)
LGL BADEM-WÜRTTEMBERG (GE)							(Urbanke & Wiedemann 2014)

* standard version, ** lite version

The differences of implementation between the NMAs are due to some factors, as Duchêne et al. (2014) explain:

- The order to tackle automatic generalisation differed per NMA. Some started with solutions to solve a particular problem while others tried to develop a production system for multiple map scales;
- Most NMAs adopted a ladder approach (deriving smaller-map scales from the closest larger map scale), but Ordnance Survey uses a star-approach;
- The level of automation and the allied need for manual postproduction defines if the system should enable incremental updates. This is also connected to the view on quality taken by the NMA.

1.4.2 Introduction of the case study

This thesis studies generalisation for manmade water networks and uses the implementation of Dutch Kadaster as case study. Dutch Kadaster replaced the interactive, manual generalisation production process of the topographic map, scale 1:50.000, with a fully automated production process at the start of 2013. The novelty is not the introduction of generalisation automation, as previous paragraphs have already shown, but the uniqueness is in the thoroughness of the implementation and in replacing the existing, traditional map (in contrast to the introduction of the new OS MasterMap™ by Ordnance Survey (Revell et al. 2011; Regnauld 2014)). The Dutch automation encompasses the whole generalisation process, while adopting a new paradigm in cartographic requirements and generalisation quality assessment. (Stoter, Post, et al. 2014) This was achieved by two major changes in existing paradigms.

In contrast to traditional interactive generalisation and most recent attempts to automate generalisation (Stoter et al. 2009), the focus was not laid upon compliance with cartographic rules. It was allowed to abandon the original cartographic requirements if necessary to achieve better results. Examples are the changed hierarchy of topographic objects (highways replace dammed dikes) and the implementation of other building generalisation requirements (i.e. detached houses are no longer treated individually and ribbon development is allowed to be portrayed as urban area). This changed view, where cartographic requirements are seen as guiding principles, is the first characteristic of the change in paradigm.

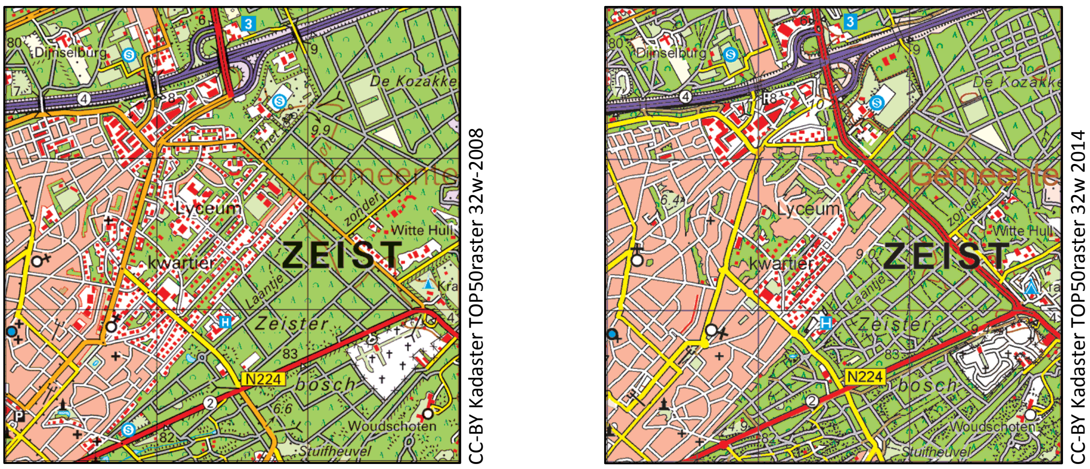


Figure 5: detached houses (left result of interactive generalisation, right result of automatic generalisation)

Figure 6: ribbon development (left result of interactive generalisation, right result of automatic generalisation)

The second characteristic is the view on quality assessment. In the past, quality was defined in compliance with the generalisation requirements. These were laid down in order of the client, the Dutch Ministry of Defence and were unchanged for about 60 years (Stoter et al. 2012). Main focus was on the military use and various cartographic decisions were dictated by the needs of the soldier.

> *To guide soldiers in the search of hiding places, a distinction was made on the map between deciduous and coniferous forest. In winter coniferous forest will keep its needles, when deciduous forest loses all its leaves. Someone looking for a shelter should take the needles for granted.*

Nowadays these kinds of map requirements seem to be fully outdated. Users and use of geo-information have changed drastically in the last ten years. Geo-information is no longer solely the domain of the army and GIS experts, but is now applied in several other domains and laymen applications. Map users are looking for maps where the content is as up-to-date as possible and are willing to give in on aesthetical and cartographic requirements. The introduction of mobile devices also changed map use and requirements dramatically.

These developments urged Kadaster to invite the map users to evaluate the results of automatic generalisation as early as possible. Instead of dictating quality, the quality standard now is set by a willingness to use the map. Maps are appreciated by serving the end users' goals. This means that a certain threshold in quality (i.e. legibility) has to be met (Bruns et al. 2010; J. E. Stoter et al. 2012), but it also changes the requirements on cartographic quality.

This new view on quality assessment gave space to experimental research in automatic generalisation. Users were asked to evaluate the provisional results, which enabled the ongoing research to target at the main interests. A design principle is that, even though errors are known in the generalisation process (Stoter et al. 2012), these are not solved afterwards by human interaction. Instead, focus is laid upon improvement of base data and sophistication of generalisation algorithms and processes.

1.5 Problem statement

Although the automated generalised maps are currently in production, there are still situations that can be further improved. One of the issues for improvement is the thinning of manmade water networks. Existing water thinning algorithms are not applicable to manmade situations, because they are biased towards nature-formed water networks. Dutch hydrography network on the other hand bears the characteristics of a cultivated, manmade network and can be better equated to a road network (Stoter et al. 2012)

For this reason, the "seemingly illogical decision" (Gould 2014, p.6) has been made to apply a road thinning algorithm for the thinning of hydrography (Punt & Watkins 2010), but results are not always convincing. Output of the generalisation process is unsatisfactory in two ways:

1. it fails in considering landscape characteristics. Differences can be observed in e.g. feature morphology (natural or artificial patterns) density and water type (canals, small ponds or rivers).

2. adjacent areas with almost equal hydrographical density are sometimes generalised totally different. Figure 7 below illustrates this by an equation of source and resulting data. (Area A is still dense, while area B is almost blank).

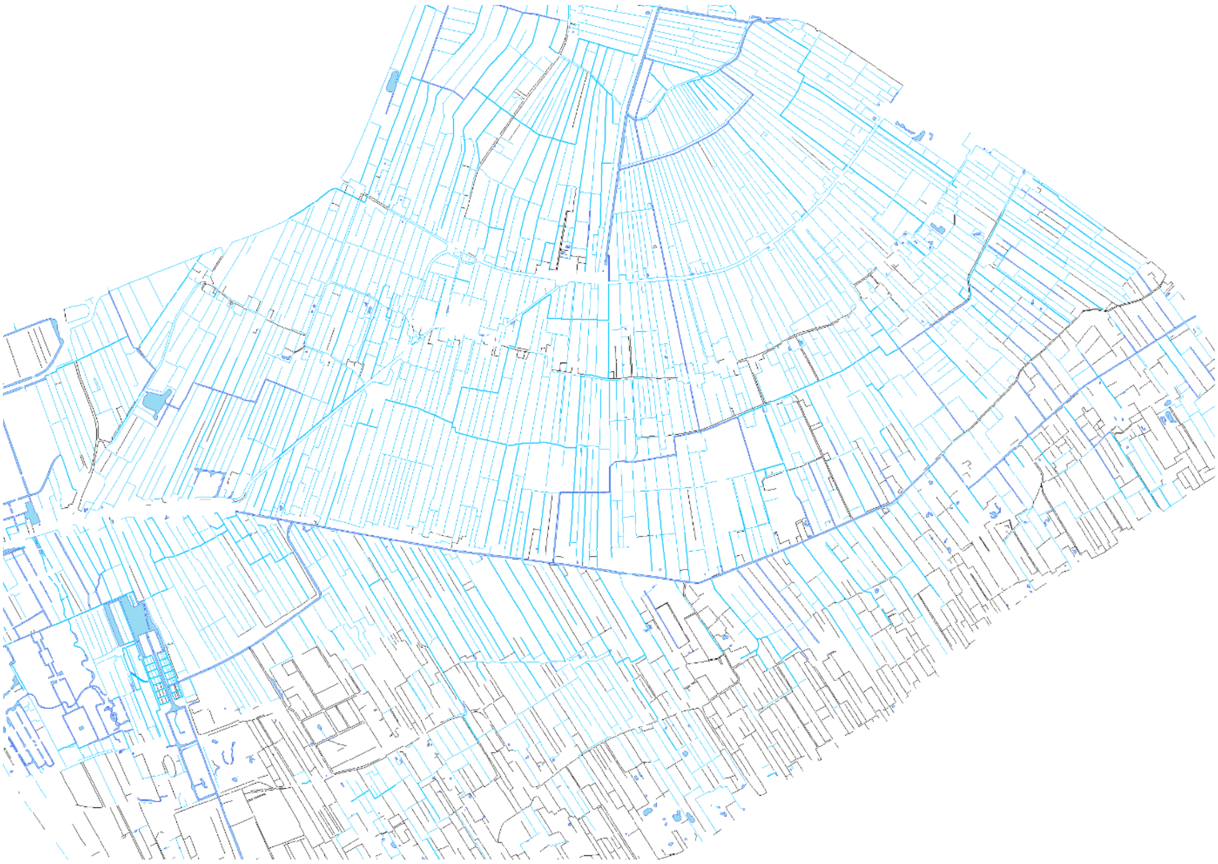


Figure 7: input data

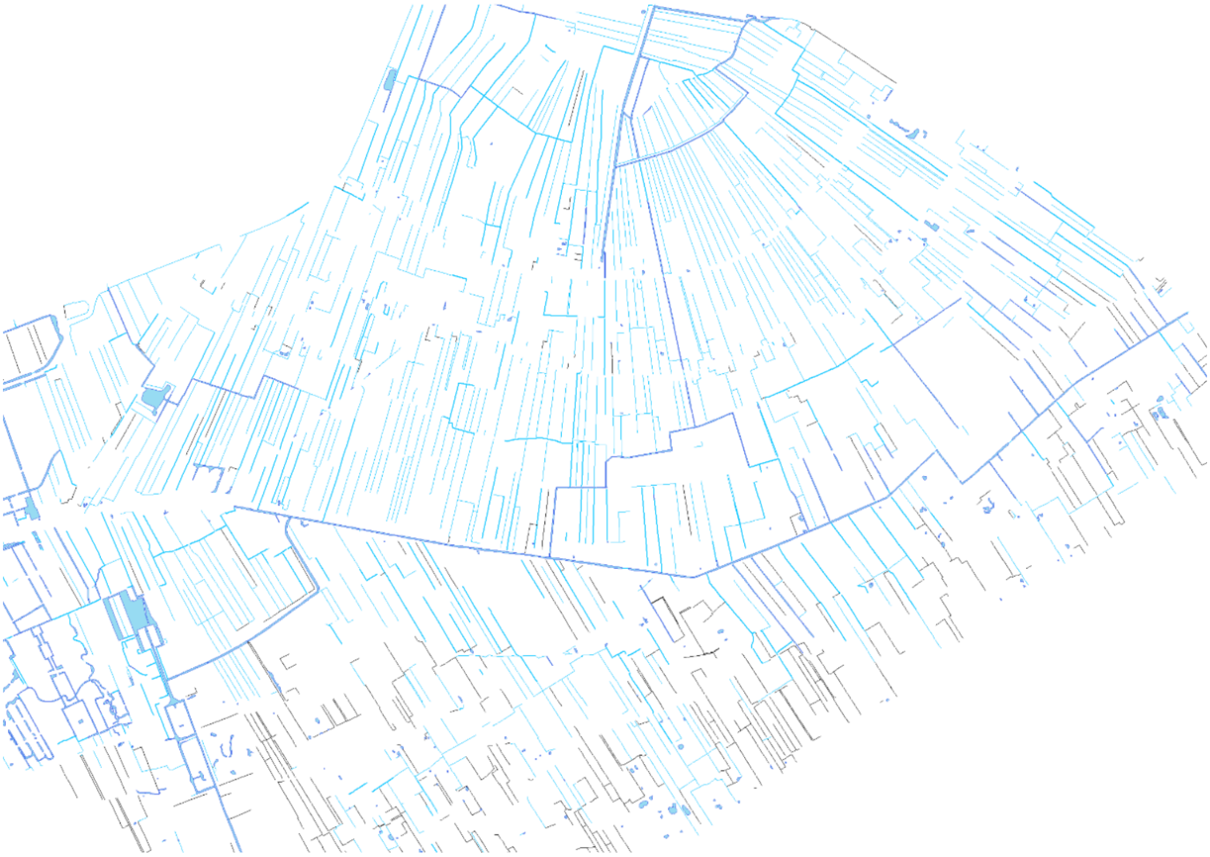


Figure 8: current thinning results of the same area

1.6 Research objective

The aims of this research project are to deliver a methodology for pruning of artificial manmade networks with regard to landscape typology and to research methods for evaluation of the quality of generalisation results. To investigate what adaptations and additions to an existing generalisation workflow are useful, the Dutch situation will act as a case study in experiments to improve the pruning of a manmade water network using different methods like data enhancement, data characteristics, data enrichment, algorithm enrichment and landscape type awareness.

1.7 Research questions

To achieve the formulated research objective, two closely related research questions have to be answered. (1) What is a suitable methodology for pruning artificial manmade network which takes landscape typology in account? And: (2) How can thinning results be evaluated objectively?

Both research questions can be broken down in to several sub questions. For the first question these can be formulated as:

- What data enhancements are suitable to establish network and improve generalisation of manmade water networks?
- Which explicit and implicit data characteristics are available or can be calculated to support efficient network thinning?

- How could these be used in generalisation of manmade water networks and what effects do they have? Which alternative network thinning algorithms are available? Are they suitable for an artificial network?
- Does context-dependent network pruning give better results and does landscape typology have a positive effect when used in network pruning? What parameters suit best?

The second research question can be broken down into two sub questions

- Which methods exist to evaluate network pruning results objectively?
- How can these be implemented into the research?

1.8 Methodology

For answering the research questions several methods are used: a literature review covering the background of generalisation approaches with a specific focus on existing thinning algorithms and the usage of contextual (landscape aware) knowledge. Existing quantitative evaluation methods are discussed and evaluated from theoretical and practical perspectives. Experiments are performed to test theory in practice and can be divided into the following:

- **data enhancement:** the waterline features in the source data are very often not suitable for network analysis. The addition of line connectors (which may be representations of real topographic elements or artificial constructions) could help to establish a water network and result in a more satisfactory generalisation.
- **exploitation of data characteristics:** Top10NL is an attribute-rich dataset which contains several explicit characteristics with regard to the water network. Besides the feature class "*waterdeel*" (water segment) which stores geometry and attributes of hydrographic features, TOP10NL also contains several other features (e.g. attribute values about main drainage and water names, position of culverts, sluices) which constitute the theme water and could provide a more satisfactory generalisation.
- **algorithm enrichment:** applying alternative network thinning algorithms.
- **landscape type awareness:** the relation of data pruning and terrain characteristics could be considered in selection of appropriate pruning algorithms and parameters.
- **evaluation:** quantitative evaluation methods are researched and developed. Qualitative evaluation is intentionally placed out of scope for two reasons: (1) because the experiments are part of a larger stack of generalisation operations the achieved results are only *intermediate* results (2) and visual evaluation requires expert knowledge.

1.9 Thesis outline

The research is divided into three parts. The first part, *Defining the scope,* contains this introductory chapter, a literature overview and an introduction to the case study. *Literature overview* focusses on three subjects. Firstly, an overview of generalisation modelling approaches is given, followed by a discussion of existing thinning algorithms especially focussing on thinning of artificially created (water) networks. Secondly, landscape typology for the Netherlands is discussed together with an overview of the use of context-dependent pruning, the third subject describes current issues and provided methods in quality assessment of generalised data.

Case study introduction, defines the use case of this research in more detail: the generalisation of water networks for the Dutch situation. It describes the datasets and context in which the research is embedded.

In the second part, *Experiments, Results and Finding* several experiments are carried out to improve the pruning of manmade water networks by individual or joint application of one or more of the following methods:

- development of location based terrain characteristic parameters to support context-dependent pruning.
- geometric preparation of base map data for better results in generalisation of manmade networks
- exploration of use of explicit water network characteristics in TOP10NL
- application of alternative water network thinning algorithms
- development and application of evaluation metrics

The experimental fieldwork has been implemented as prototype tools and results have been assessed by evaluation tools. All tools have been built in ArcGIS ModelBuilder, if necessary expanded by Python-scripts. It is conceivable that these methods will deliver differing results for different landscape types. The necessity of tailoring methods for different landscape types will be tested on several study areas and a subsequent comparison of the results of current algorithm implementation with those of alternatives will try to evaluate the methods.

The results of the alternative approaches to the different landscape types are discussed and quantified by using appropriate evaluation methods which were identified in the *Literature review* and by own developed methods. Also the results of the alternatives are discussed in comparison and coherence with one another.

The last part, *Conclusion,* contains a concise overview of the research results. It will also identify the issues for further investigation. Figure 9 gives an overview of the structure of this thesis.

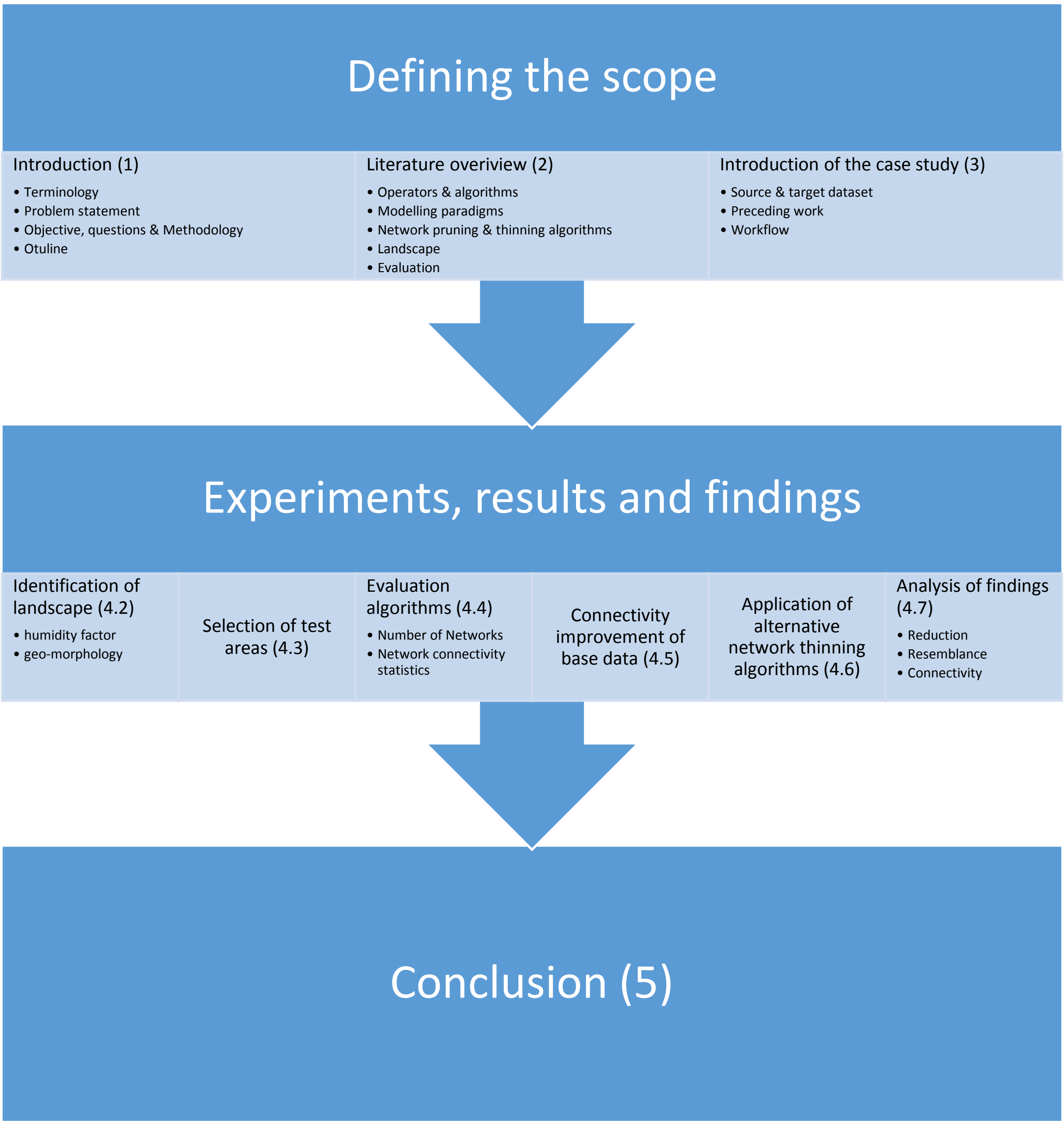


Figure 9: a graphical overview of the research process

2 Literature overview

Over time a vast amount of research has been published in the field of generalisation, network thinning and evaluation of generalisation research. Although these researches cover a wide variety of approaches, this review will treat six topics relevant to this study. These topics include the definition of operators versus algorithms, a discussion of generalisation modelling paradigms, the need for network pruning, a concise overview of network pruning algorithms, the relevance of landscape in generalisation and a section which discusses the difficulties in and proposed methodologies for evaluation of generalisation results. While the original researches focussed on a broad range of contexts and applications, this chapter discusses these topics within the context of network pruning of artificial networks.

2.1 Operators and algorithms

As use of operators and implementation of algorithms is common to any generalisation modelling paradigm, it is important to state the difference between these two concepts beforehand. Stanislawski, Buttenfield, et al. (2014, p.158) define a generalisation operator " … as a generic descriptor for the type of spatial or attribute modification to be achieved on some set of geospatial data." Typical examples of generalisation operators are simplification, enlargement, displacement, merging, selection, symbolisation and enhancement. (This list is not exhaustive and in the literature sometimes synonyms are used).

Attempts have been made to class generalisation operators , i.e. (Kraak 2010, pp.99–100), but often these classifications do not fit usage in an automated process. To overcome this difficulty, Stanislawski et al. (2014) classify operators in three classes: pre-processing, quantity reduction and aesthetics (visual quality) as outlined in the figure below:

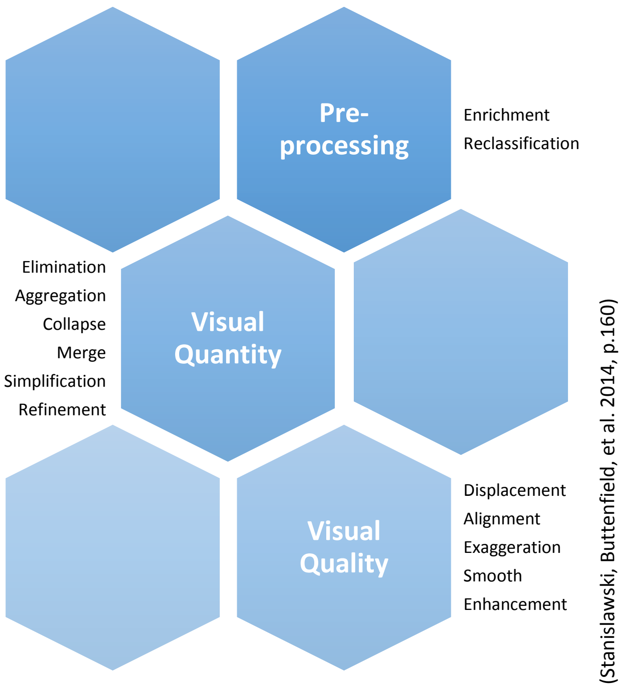


Figure 10: classification of vector generalisation operators

This study focusses on the operator which refers to "…the removal of one or more features without replacement." (Stanislawski, Buttenfield, et al. 2014, p.161) Several synonyms are used for this operator: *elimination*, *selection*, *class selection*, *extraction*, *thinning*, or *pruning* (see Stanislawski, Buttenfield, et al. (2014), for the appropriate references).

However, to efficiently and satisfactorily thin a network or more in general to apply some kind of generalisation, different operators are involved, often within a certain sequence. The actual implementation of a (series of) operator (-s) is defined as "algorithm":

> "An algorithm refers to a specific method by which one or more operators are implemented, and several algorithms usually exist for the same operator." (Stanislawski, Buttenfield, et al. 2014, p.158)

> "An algorithm may implement one generalisation operator (e.g. building amalgamation) or a combination of operators (e.g. road simplification and smoothing)." (Gould 2014, p.222).

To summarize the differences: an operator describes at a *conceptual* level the generalisation action to be performed, whether an algorithm is the practical tailor-made *implementation* of a (series of) operator (-s) in a certain context.

2.2 Modelling paradigms

Results of early research in automatic generalisation were atomic algorithms which only solved individual generalisation problems. Examples are line generalisation, e.g. the Douglas-Peucker algorithm (Douglas & Peucker 1973) and displacement, e.g. Jäger (1991). In the early 1980s the need arose for an overarching approach which would chain all the atomic units together. Four approaches to generalisation modelling which are commonly distinguished (Harrie & Weibel 2007) developed subsequently: batch processing, condition-action modelling, human interaction modelling and constraint-based modelling (which includes agent based modelling).

The first attempts to automate generalisation were based on **batch processing,** which was used to chain individual generalisation operations in a fixed sequence (Regnauld et al. 2014). All algorithms, parameters and input data were prerequisites to run and consequently interaction was impossible during runtime. A change in parameter values required to rerun the whole process. The system was lacking flexibility and conditional procedures.

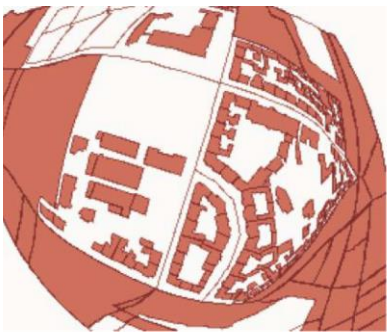
Condition-action modelling can be regarded as the successor of batch processing. This is a two-fold approach which was developed and implemented in the late 1980s. Starting with a lot of pre-defined conditions stored in a rule-base, the first step is to recognize which condition is applicable to each individual structure in the source data, which results in 'structural knowledge'. After identification, an appropriate action (based on the identified condition) will be triggered for each identified structure to generalise the source data. While the condition-action modelling approach has got off to a flying start, several problems arose during implementation. The formalisation of cartographic rules appears to be very difficult. Furthermore the necessity of a multitude of rules for maintaining mutual relations of map objects appears to be overwhelming. And in the end, the system is unable to evaluate the results of its operators and adapt its processes consequently.

This complexity and corresponding issues led some in the early 1990s to develop **human interaction modelling**, a hybrid approach where both human and computer cooperate and share responsibility in decision-making on complex cartographical problems. The computer can perform the tasks which can be formalised, but also aids the human in making complex decisions. The human guides the computer in providing the correct application order of algorithms and in evaluating intermediate results. Weibel suggested the term *amplified intelligence*: "…human intelligence is amplified by the processing power of the computer, while the limited capabilities of the computer for holistic reasoning as well as visual perception and quality evaluation of the generalisation results are amplified by the presence of human intelligence." (Harrie & Weibel 2007, p.71). A new approach arose in the mid-1990s due to insufficient implementation in commercial systems, the ongoing need for expert users and a lacking decline in manual labour.

Constraint-based modelling shares characteristics with condition-action based modelling as it uses a rule-database to achieve generalisation results. However, the function of rules in both approaches is different: instead of defining which actions to perform, focus is laid upon defining the environment and the result of the actions which should be achieved. Constraints limit the bandwidth of acceptable results. The process to achieve these results is far more flexible allowing the system to alter the order of operations during more iterations and choosing the most optimal solution (balancing time and results). Approaches using this modelling paradigm are agent and multi-agent systems which are regarded as the most potential variant of constrained based modelling (Harrie & Weibel 2007).

Agent systems are capable of handling complex situations by using meso-agents (overarching operators), can (potentially) deal with a complete set of generalisation operators and are suited to integrate other modelling techniques. Their success, however, depends heavily on the accuracy of constraint definition and the measurement precision of evaluation rules. Other complicating factors are the difficulty to provide a meaningful clustering of agents into a meso-agent, and issues in achieving required formalisation of procedural knowledge to guide generalisation algorithms and parameters. A lot of research to improve and overcome the identified issues has been and is still being undertaken, e.g. (Lamy et al. 1999; Barrault et al. 2001; Galanda 2003; Revell et al. 2005; Duchêne et al. 2012). Implementations can be found at Ordnance Survey UK in the AGENT-system (Lamy et al. 1999) and at IGN-France in the CartaCOM-system (Duchêne et al. 2012).

At Delft University a fifth paradigm is being researched: **Vario-scale modelling** (van Oosterom et al. 2008; Meijers 2011; van Oosterom & Meijers 2011)**.** This paradigm aims to provide a structure for continuous generalisation of real world objects thus enabling smooth zoom actions and consistent mixed-scale representations, see Figure 11 below. It also aims at solving the problem of data redundancy and inconsistency other approaches for multi-scale mapping are facing. Areas for application are generalisation of multi-scale maps or generation of transition areas for 3D visualisations. The current state of the research is too immature for implementation in production lines.



Van Oosterom & Meijers (2011)

Figure 11: an example of a mixed-scale map

2.3 Network pruning

In all modelling approaches special attention is paid to incorporate network pruning. A lot of research on network pruning algorithms has previously been undertaken, which can be divided in two categories: road network thinning, e.g. (Thomson & Richardson 1999; Edwardes & Mackaness 2000; Thomson & Brooks 2000; Travanca Lopes & Catalão 2002; Jiang & Claramunt 2004a; Jiang & Claramunt 2004b; Chaudhry & Mackaness 2005; Heinzle et al. 2005; Han & Qiao 2010; Luan & Yang 2010; Wang 2011; Li & Zhou 2012; Altena et al. 2013; Benz & Weibel 2013; Weiss & Weibel 2013) and hydrographic network thinning, e.g. (Stanislawski 2008a; Stanislawski 2008b; Stanislawski 2009; Stansilawski et al. 2009; Buttenfield 2010; Buttenfield et al. 2010; Anderson-Tarver et al. 2011; Savino 2014).

However, categorisation of network pruning algorithms based on theme (water or road) does not always reflect reality, which is illustrated in the figures below:

Figure 12: the Dutch water network

(O'Beirne n.d.)

Figure 13: a braided river

(Campbell n.d.)

Figure 14: street network in San José

(The Lotus Forums 2014)

Figure 15: the Gotthard pass

Geometrical and topological structures are more suitable to discern the appropriate thinning algorithm, as was evidenced by Li & Zhou (2012) and experienced in the Dutch implementation of a road thinning algorithm to prune the water network, cf. (Bruns et al. 2010).

2.4 Existing network pruning algorithms

Pruning of a network is often performed using a stack of methods, within a logical order: the identification of importance of elements within the network, the reduction of elements based on the established importance, identification of graphically conflicting features and subsequent elimination. The indication of importance of the individual objects within the network is often taken as starting point. Despite many variations in detail, this first step can be divided into three main categories according to Liu et al. (2010): *semantic-based*, *graph-based* and *stroke-based*. These will be discussed below. In addition the *mesh-based* and the *stroke-mesh-based* approaches will be discussed also. These approaches are the results of some experiments in extending or combining methods to overcome some limitations.

2.4.1 Semantic-based

Semantic-based selection uses attribute information to provide information for selective omission of network features. Considering a hydrographical network, examples are type of water (in hierarchical order: trench, watercourse 1, watercourse 2), width category of objects, part of main drainage, etc. The main drawback of this approach is the neglecting of topological and geometrical structures (Liu et al. 2010). This is illustrated in Figure 16, where a straightforward selection is performed on the highway network omitting objects with the attribute “on exit”.

Figure 16: an incomplete road network - one of the problems a semantic selection can lead to

2.4.2 Graph-based

Mackaness & Beard (1993) applied graph-theory to the field of generalisation. Graph-based selection uses the visual representation of spatial relationships between objects to establish their importance. The main focus is on the connectivity of elements individually, and mathematical concepts (shortest path, minimum-spanning-tree and degree of centrality) are used to determine the individual ranking of objects within a network, which has been applied to recognize a city centre. Attempts to use the approach for discovering graphical patterns in networks (Heinzle et al. 2005) were promising, but did not compensate another issue: ignoring thematic and geometric characteristics of the objects which is seen as a major limitation of the method (Liu et al. 2010).

2.4.3 Stroke-based

Stroke-based selection is a two-step method to establish the importance of the individual network objects. It derives its name from the idea of a ‘stroke’ which Thomson and Richardson (1999) define as “… a curvilinear segment that can be drawn in one smooth movement and without a dramatic change in style”. The underlying assumption is “…traffic routes are built as curvature-poor as possible" (Heinzle et al. 2005, p.3). The main principles are 'good continuation' and 'similarity of characteristics'.

A network structure is broken down into topological edges and nodes. The relative importance of each edge within the network is identified by aggregating the individual elements into larger ones based on one or more shared characteristics and predefined rules (i.e. which characteristics are allowed to be dissolved with higher order characteristics). The second step is performing a selection, using the identified hierarchy and a predefined threshold.

The method is being criticised, i.e. by Liu et al. (2010), for neglecting overall and statistical information, which can lead to biased outcomes. This disadvantage must be solved will this method be of use in generalising networks in different landscape regions. Research to solve this problem has focussed on the selection of shared characteristics for stroke-construction. These can be identified on basis of geometry, theme or a hybrid combination of both.

In *geometric-stroke-construction* three variants can be distinguished (Jiang et al. 2008): the *self-fit*, the *self-best-fit* and the *every-best-fit*. Within the *self-fit* approach, any edge which satisfies a predefined threshold angle, is a candidate for a stroke. The algorithm chooses at random which candidate to use. To overcome this unpredictability the *self-best-fit* approach has been developed. Here only the edge which satisfies the threshold value best (i.e. has the smallest deflection angle) will be used to construct a stroke. The third variant, *every-best-fit*, expanded this variant further. In this approach, *every* edge which satisfies the threshold is used to construct strokes, which could result in one edge being part of multiple strokes. This kind of information can be used to identify a hierarchy for edges within the network, which is also based on the relative importance of the edges.

Another approach is the *thematic-stroke-construction,* using a common attribute value to identify potential strokes (Zhou & Li 2012). This has been applied by Jiang & Claramunt (2004b) using street names, but at least three problems are identified with this method: (1) attribute values are not complete within the source data; (2) attribute values are not always consistent, i.e. when indicating sub-units of an element; (3) situations can occur where there are multiple candidates for a stroke, each with a corresponding attribute value.

The *hybrid-stroke-construction* approach (Zhou & Li 2012) tries to use the advantages of the geometric approach to cover up the weakness of the thematic and vice versa. By combining the results of both approaches better choices can be made in constructing strokes, since information is available about geometric relations, but also about thematic relations.

2.4.4 Mesh-based

The mesh-based approach (Edwardes & Mackaness 2000; Chen et al. 2009) focusses on density. It centres on the concept of a *mesh,* an area fully enclosed by edges (i.e. roads). The density of a theme (i.e. a road or a waterway network) can be calculated by the construction of meshes out of the individual objects of the theme. The density D for each mesh individually is computed by using the flowing formula:

$$D = \frac{P}{A}$$

where *P* is the perimeter and *A* is the area of a mesh (Li & Zhou 2012). Subsequently, the meshes which do not satisfy a predefined threshold are merged with one of their neighbour and the original edge feature must be removed. By iterating the process, all meshes will finally meet the threshold.

A threshold in the range of 0.4 mm to 0.7 mm has been proposed in previous research (Zhou & Li 2012), but experiments can help in establishing this value. Chen et al. (2009) provided a formula to calculate the mesh density threshold (Tm):

$$Tm = \frac{4}{StLm\left(1 - \frac{Ss}{St}\right)}$$

where Ss and St are the scale factor of respectively the source and the target map and Lm is the size in map distance (diameter or length) of the smallest visible object on a map. While the scale factors are known variables, the size of the smallest visible object (Lm) can vary depending on the map specifications.

The concept of a mesh-based pruning approach is illustrated for the graph in Figure 17. The graph consists of 18 edges out of which 6 complete meshes can be constructed (A to F). Mesh G cannot be constructed because of missing edges.

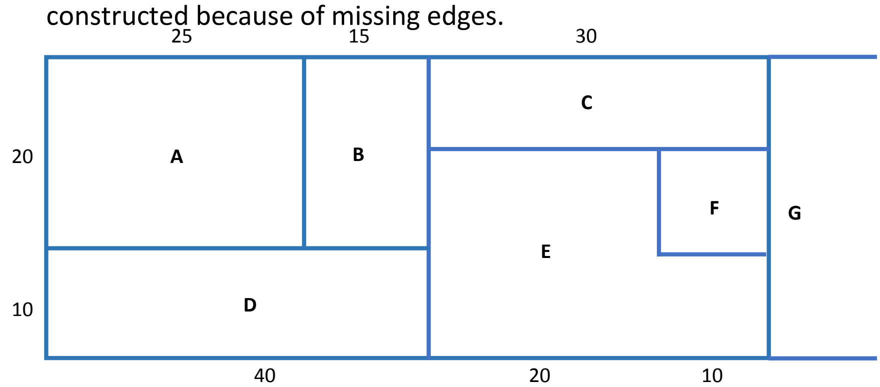


Figure 17: initial graph for mesh-based thinning

In Figure 18 for each mesh the area is calculated and Figure 19 shows the accumulated length of edges per mesh.

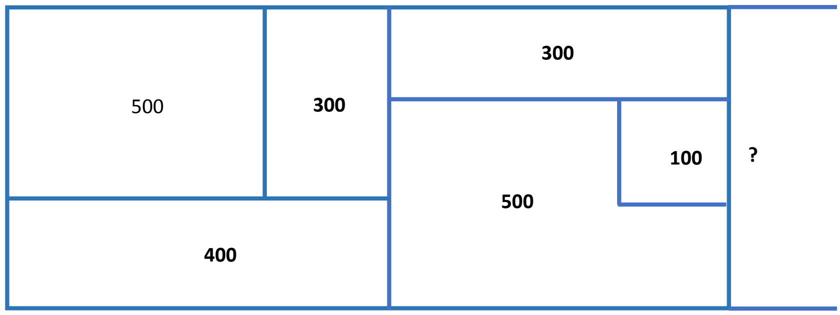


Figure 18: calculated area for each mesh in the initial graph

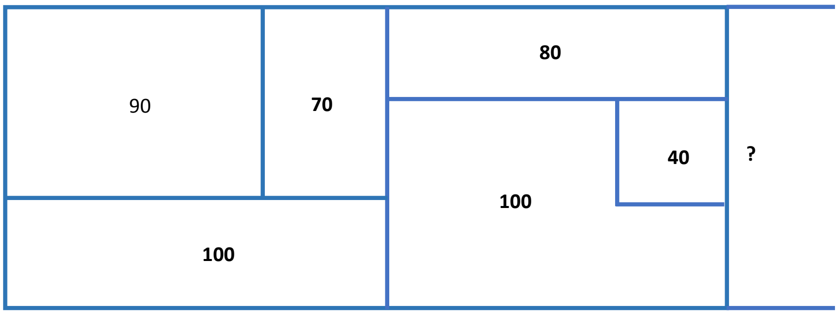


Figure 19: accumulated edge length for each mesh in the initial graph

Based on these figures the density has been calculated for each mesh using the formula above as provided by Li & Zhou (2012). For the graph in Figure 17 this results in the figure in Table 2

Table 2: computed densities for meshes in Figure 17

Mesh	Calculation	Density (D)
A	90 / 500	0.18
B	70 /300	0.23
C	80 /300	0.27
D	100 / 400	0.25
E	100 / 500	0.2
F	40 / 100	0.4
G	unknown	unknown

These density values can be used to determine which meshes have to be eliminated based on a provided threshold. The threshold for the generalisation of 10k to 50k data can be calculated (using the formula suggested by Chen et al. (2009)) assuming a diameter or length of a the smallest visible object of 20 meters (which is the equivalent of 0.4 mm on a 50k map):

$$0.25 = \frac{4}{50.000 * (0.4 * 0.001) * \left(1 - \frac{10.000}{50.000}\right)}$$

This means each mesh with a density larger than 0.25 has to be eliminated in an iterative action, starting with the smallest mesh which has to be added to a larger neighbour. Two candidates are available to add the small polygon to: addition to the largest neighbouring polygon or addition to the polygon with the longest shared side.

In the example below the choice was made for the largest shared side. So first, mesh F will be selected (having the highest density value) and will be added to mesh E. Then the densities are recalculated and the mesh with the highest density value will be added to its neighbour. Since the new mesh E' (green n the picture below) will have a density value of 0.167, the mesh with the highest density value will now be mesh C (0.27). On basis of the longest shared edge, mesh C will be added to the newly created mesh E' resulting in mesh E'' (orange in the figure below) with a density value of 120 / 900 = 0.13. The iteration for this example will stop after two times, as all meshes will have values which do not exceed the calculated threshold.

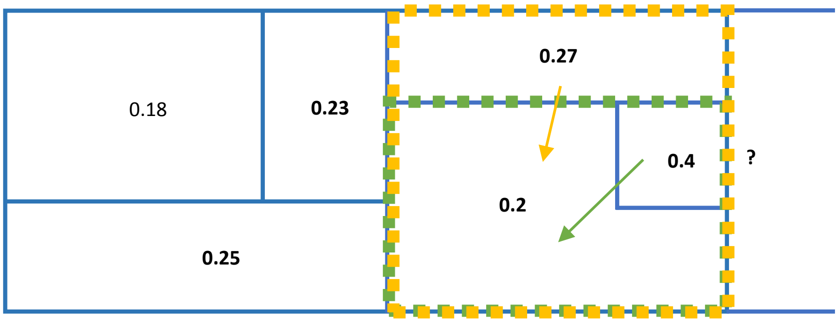


Figure 20: calculated density (D) for each mesh

A mesh-based approach delivers usable results for a simple graph. However, there are some shortcomings in this algorithm. The algorithm excludes mesh G in the graph above from the analysis and subsequent pruning. While it is clear that mesh creation is impossible for G, this affects pruning. Another shortcoming of the algorithm is revealed when it is applied to real-world situations with dead-end roads. This is illustrated in Figure 21 when a dead-end road is added to the previous graph and the mesh-based approach is applied.

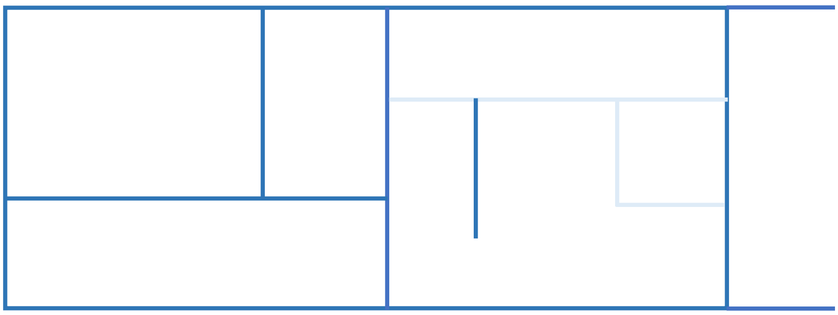


Figure 21: shortcoming of a mesh-based approach to handle dead-end roads

This results in a reduction of network connectivity and the undesired isolation of an edge.

2.4.5 Extensions of algorithms

To overcome the limitations of the aforementioned methods, there have been some experiments which try to extend or combine them. One of these is the **stroke-mesh based approach** as first proposed by Li & Zhou (2012) and experimentally tested by Benz & Weibel (2013) on medium-scale (50k) data of swisstopo. Benz & Weibel established two strengths of this algorithm: details of a network are handled carefully (important for medium-scale generalisation) and the general connectivity of the network is being maintained.

The stroke-mesh approach rests on the concept of areal and linear segments (Benz & Weibel 2013). Areal segments coincide with at least one mesh edge, while linear segments do not share sides with edge meshes at all, see Figure 22.

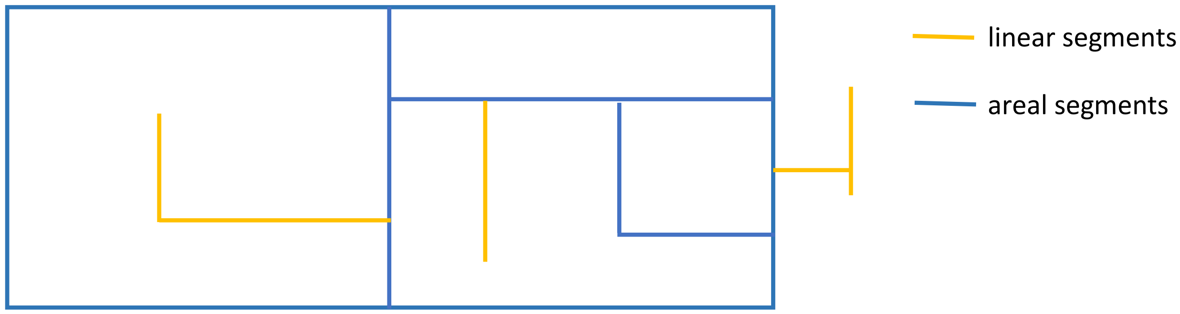


Figure 22: areal versus linear segments

Li & Zhou (2012) demonstrated by experimental research mesh-based thinning suits areal segments best, while stroke-based thinning better applies to linear segments. This led them to combining both methods in a hybrid approach which has alternative hierarchy establishment paths dependent on segment-type. Results of both pathways are incorporated in a later stage and in a new iteration a new selection is applied interconnecting both linear and areal segments. The initial algorithm has been extended by Benz & Weibel (2013) to obtain better results. This only has been tested for four areas and was evaluated by swisstopo experts. Results are regarded as promising, but pre-mature: more research is necessary on other test areas and the applicable scale range has to be established by experimental research.

2.4.6 Thin Road Networks

In the geoprocessing toolbox of the commercial software package ArcGIS an algorithm is available which is aimed at the thinning of networks: Thin Road Network. Only a functional description of the algorithm is available for end-users, which can be summarized as follows:

A "…simplified road collection is determined by feature significance, importance, and density. Segments that participate in very long itineraries across the extent of the data are more significant than those required only for local travel. Road classification, or importance, is specified by the Hierarchy Field parameter. The density of the resulting street network is determined by the Minimum Length parameter, which corresponds to the shortest segment that is visually sensible to show at scale." (ArcGIS Help 10.2 2014)

An understanding of the optimisation concept is important to appreciate the results of the algorithm. "The Thin Road Network tool … is built on an underlying generic optimisation engine (Monnot et al. 2007a; Monnot et al. 2007b). It makes use of an in-memory cache of explicit topology, so as to be able to efficiently impose topological constraints (e.g. retaining connectivity of roads). The optimisation engine has the task of finding a 'good enough' solution in reasonable time to a combinatorial problem space, which if explored exhaustively would take an enormously long time to traverse. It does this by trying more drastic solutions early on, and as time passes, refines more local situations. […] The internals of this optimiser-based strategy are hidden from the user, as the functionality is exposed as an ArcGIS geoprocessing tool with a small number of user-settable parameters […]." (Mr. P. Hardy 2014, pers. comm., 18 November)

2.5 Landscape

Previous paragraphs have shown several pruning algorithms. Some are more applicable for mesh structured networks, while others are more applicable to stroke-based networks, while in reality both networks exist in landscape. But what exactly is meant by landscape?

In this research the definition of Pungetti is taken which sees landscape as a '…dynamic process developing on the visible earth surface, resulting from the interaction between abiotic, biotic and human factors which vary according to site and time.' (Makhzoumi 1999, p.6) Important in this definition is the interaction between natural and artificial developments which applies to Dutch landscape. Differences in landscape can be categorised as landscape types, compare e.g. Farjon et al. (2001), and are of influence in generalisation. Figure 12 to Figure 15 already illustrated how the difference in structure of geographical objects is due to scenery. Brewer et al. advise that "...generalisation components must be varied with landscape regime." This variance can be applied to algorithms, parameters and the sequence of processing steps. "Pruning is necessary in some but not all landscape regimes." (Brewer et al. 2009, pp.5–6).

With regard to network thinning, at least two approaches in establishing landscape characteristics are conceivable. The first approach is deductive by superimposing already defined landscape typologies on individual features, while the second approach is inductive and uses computation of feature characteristics like average feature morphology and density to identify the landscape type.

In a deductive approach issues can be encountered since existing landscape typologies differ. They are designed for utilisation in a specific context (compare e.g. the classifications of Farjon et al. (2001, pp.77–78) and Cultural Heritage Agency of the Netherlands (CHA-NL) (2014). The first is designed for usage in a discussion on landscape development balancing conservation of cultural heritage and stimulation of economic development and has 54 classes, while the second aims at "listing, preserving, sustainably developing and providing access to the most valuable heritage in our country" (Cultural Heritage Agency of the Netherlands 2014) and consists of 9 main types which together hold 19 subtypes.

The inductive approach to establish landscape typology can be based on the computation of geometrical characteristics of features and the density of features. It is common to distinguish landscape types on basis of characteristics like humidity and density (ref). This can be established by computing the ratio of water features compared to the overall area. Another method is based on the appearance of features: what is the morphology of individual features and, from a network perspective, how is the connectedness of the network, see 4.2.2.

Instead of starting a quest for the most optimal landscape classification with regard to network thinning or starting from scratch with a massive computation to establish landscape areas, a hybrid approach to landscape is applied in this research. The main classification of the Cultural Heritage Agency of the Netherlands is taken as a starting point. This classification is evaluated for usability with the question in mind: "is the provided landscape typology suitable for identification of distinct landscape areas to orchestrate a generalisation workflow?" To answer this question, three tests are performed: for each landscape type (1) the feature morphology, (2) density and (3) number of networks are calculated for all water features. This information is used to decide if the Landscape typology of CHA-NL is fit to tailor the application of pruning algorithms.

2.6 Evaluation

Generalisation results in quality change of data (Muller 1991), but often these changes were intentional, as it makes the results more fit for use at a smaller scale. "[…] the reference data and the data evaluated are at different scales […and…] corresponding relations identifying objects that represent the same object in reality." (Stoter, Zhang, et al. 2014, p.272) Therefore, a proper methodological framework is needed to judge the successfulness of a generalisation procedure. This framework is commonly referred to as *Evaluation* which according to Stoter, Zhang et al. (2014) can be defined as:

> "… the process of examining and checking whether the desired characteristics of the resulting data are satisfactory for a given task." (Stoter, Zhang, et al. 2014, p.260)

The quality of a generalisation process from a data perspective can be established by two approaches: the first focusses on the question whether results can still be regarded as a faithful presentation of the *source data*, while the second approach evaluates the *results* based on the legibility of the resulting map.

From a methodological perspective evaluation of generalisation results can be qualitative or quantitative. Visual inspection is a form of quantitative evaluation and entails "… confirmation of similar content and similar densities of hydrographic detail compared to existing topographic maps […] [and] judging vertical integration of generalised hydrography with other layers, such as transportation and terrain […]." (Brewer et al. 2009, p.9) To enable visual evaluation, care has to be paid to map design to provide an unbiased representation of generalisation results. However, the results of such an evaluation method are susceptible to the drawbacks of every qualitative study: subjectivity caused by a too small sample group, unwarranted generalisation of result findings, misinterpretation or bias. Both perspectives can be integrated to distinguish an evaluation methodology, which results in the matrix below:

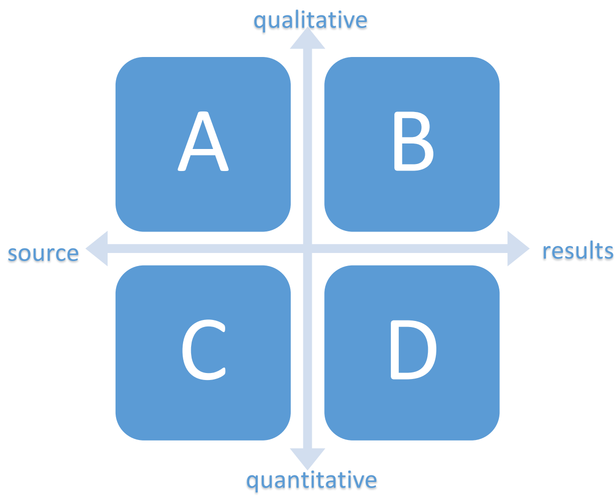


Figure 23: decision matrix for evaluation

2.6.1 Evaluation metrics

To overcome these difficulties and to enable objective, quantitative evaluation, a few metrics have been proposed for evaluating network pruning algorithms. These can be divided in two categories:

correspondence coefficients and network statistics. The paragraphs below will discuss each subsequently.

2.6.1.1 Coefficient of Correspondence

The coefficient of correspondence was first proposed by Stanislawski et al. (2010) to assess the amount of network thinning and has been implemented by Buttenfield et al (2013) as a raster method.

The method compares the results of thinning operations with an existing benchmark dataset and provides statistics to evaluate the consistency of pruning and generalisation operations. Two variants are proposed, the Coefficient of Line Correspondence (CLC) and the Coefficient of Area Correspondence (CAC). Three statistics are needed to calculate the coefficient: *conflations*, i.e. features common in results and benchmark, *omissions*, i.e. features only present in the benchmark and *commissions*, i.e. features only present in results. The summed attributes for CLC considers feature length, for CAC they consider feature areas.

$$\frac{\sum conflations}{\sum conflations + \sum(omissions\ +\ commissions)}$$

The advantage of this metric is that it provides a statistic to objectively compare different results, but for use in generalisation this metric could have a major limitation which is related to choosing a fitting benchmark dataset. When comparing results of an automatic procedure with those of a manual procedure, the problem could arise that displacement has been applied in the manual set where the automatic procedure is not displaced at all. Another complicating issue to determine possible causes for differences between results and benchmark is the difference in actuality between the datasets. Datasets are a snapshot of reality at a given time.

For this study the use of correspondence coefficients is limited to a mutual equation of automatically derived results of different algorithms, where the results of the existing implemented procedure can serve as a benchmark dataset.

2.6.1.2 Network Statistics

Another approach to measure quality of thinning can be found in the application of network statistics. “In a network database, linear features are linked together topologically.”(Wong 2005, p.304). Two approaches to assess network qualities can be distinguished. The first approach assesses the connectivity of an entire network, where the second approach assesses the accessibility of network elements. The first approach and can be divided in two methods: the assessment of the connectivity level and the determination of the capacity to support circuits by a network. Both will be researched in the following paragraphs, but first the concept of topology will be explained.

2.6.1.2.1 Assessing network connectivity level

The way segments are linked within a network and the relationships of these linkages to one another are described in terms of topology. Fundamental to assessing the connectivity of a network is *planar graph theory*. A graph is a graphical abstraction of a topological network which can be deconstructed in *edges* and *nodes* (Rodrigue 2013). An *edge* is a linear segment which is defined by the two nodes (vertices) at either end of the linear feature (Wong 2005). In a *planar* graph on each intersection of crossing edges a node is obliged. Figure 24 below shows a conceptual representation of a “real-world” structure and its topological equivalent as graph with edges and nodes.

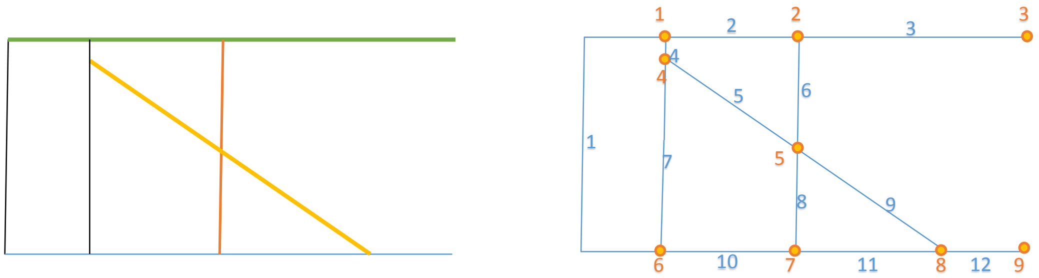


Figure 24: conceptual representation of a “real world” network (left) and a topological representation with edges and nodes (right)

The topological representation makes it clear that this network consists of 12 edges and 9 nodes. This information can be used to calculate network connectivity statistics. Wong (2005) proposes two formulas. The first is to calculate the level of connectivity and builds on the idea that every network consists of a number of nodes and a number of edges. Given a fixed number of nodes, a higher number of edges increases the connectivity level of a network. The number of edges, however, should meet a certain minimum amount (e_{min}) to connect every node within the fixed number. To assess the level of connectivity of a network, the γ-index can be calculated. This is a fraction representing the relation between the actual (e) and the maximal possible (e_{max}) amount of edges in the network.

To calculate e_{min} (the minimum amount of edges needed to connect all nodes) the formula

$$e_{min} = v - 1$$

can be used, where v is the number of nodes in the network. For the network in Figure 24, the minimal number of edges should be 8, since e_{min} = 9 – 1. Three edges can be removed from the network still keeping the network connectivity intact. However, this cannot be a random combination of edges, since it is not allowed to split the network in two separate networks.

Another important statistic in assessing an existing network is the maximal connectivity index. This index can be used to calculate the maximum possible number of edges for a fixed number of nodes. This statistic presupposes a *planar graph topology,* i.e. edges are not allowed to cross or intersect each other and can be expressed as

$$e_{max} = 3(v - 2)$$

where v is the fixed number of nodes within a network. The e_{max} for Figure 24 is 21 edges, since 3 x (9-2) = 21. This is demonstrated in Figure 25 below, where a maximum number of 9 edges can be added until violating the rule of *planar graph topology*:

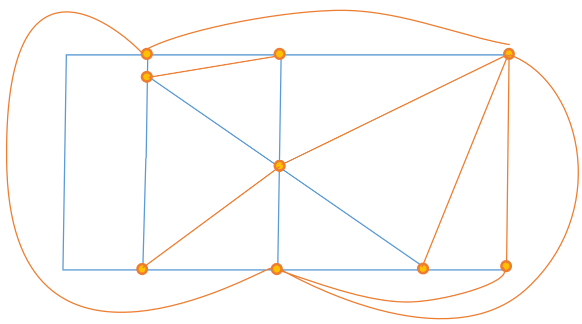


Figure 25: an example of maximum number of possible edges based on *planar graph topology*

e_{max} does not prescribe which connections have to be made to maximize the connectivity of the network, but it provides usable information for establishment of the connectivity level of a network which can be calculated by the γ-index:

$$\gamma = \frac{e}{e_{max}}$$

where e is the actual number of edges in the network. The γ-index for the network in Figure 24 is 9 / 21 = 0.42857.

2.6.1.2.2 Capacity of network circuits

Another important feature of network structures is the capacity of circuits. A circuit can be defined "as a closed loop along a network […] [where] the beginning node of the loop is also the ending node" (Wong 2005, p.308). Circuits are important indicators of alternative routes within a network. They will not appear in minimally connected networks, but the addition of an edge to a minimally connected network results in the appearance of a circuit. To calculate the number of circuits within a network a subtraction of the minimal number of edges from the actual number of edges has to be calculated. In simplified form, the formula (Wong 2005, p.308) can be written as:

$$e - v + 1$$

Equivalent to the lines of reasoning for establishing the γ-index, an index to equate the numbers of circuits of networks can be calculated, which is known as the α-index. First, the number of maximum possible circuits has to be calculated by

$$2v - 5$$

This formula still obeys the rules of planar graph theory: prohibiting crossing or intersection of edges. The simplest form of a two-dimensional area is a triangle which consists of 3 nodes. This results in 1 circuit (2 x 3 – 5 = 1). These two figures now can be used to calculate the α-index, which relates the actual number of circuits to the possible number of circuits:

$$\alpha = \frac{e - v + 1}{2v - 5}$$

2.6.1.2.3 Application of γ- and α-indices in this study

Both the γ- and the α- index are useful to equate the quality of networks (Wong 2005) and will be used in this study to compare network connectivity of input data with output data and to differentiate the outcomes of several algorithms.

An implementation of connectivity statistics has been made by the author in ArcGIS ModelBuilder and can be found in paragraph 4.4. This implementation will be used to evaluate the generalisation operations.

3 Case study introduction

The previous chapters indicated network pruning is part of a whole system of generalisation operators. These operators are usually part of a sequence to fulfil a generalisation task. Results of individual operators must be evaluated in coherence with each other. Besides these coherence, also the dependence on input data as well as the quality specifications of the desired output must be taken into account.

This chapter briefly discusses two subjects which are of importance for setting the experimenting context: first the specifications of the input and output datasets. The input dataset is Top10NL, an object oriented geodatabase. The target dataset is TOP50, a dataset aimed at cartographical use. The second topic is the generalisation workflow implemented at Kadaster, which is of importance to pinpoint the place of the pruning algorithms in the implemented sequence.

3.1 Top10NL: the source dataset

In the Netherlands, a system of key registers is regulated by Law. One of these key registers is the *Basis Registratie Topografie* and contains since 2008 a midscale topographic dataset at scale 1:10.000, TOP10NL. This is the standard base dataset for the appropriate scale with obligatory use for governmental parties.

The history of TOP10NL goes back to the start of this century, when the Dutch Mapping Agency (then known as *Topografische Dienst Nederland*) started a joint research project together with the universities of Delft, Wageningen and Enschede on object oriented databases. In this research technology, user-aspects and data-models were considered while establishing new product specifications. Finally TOP10NL would substitute its predecessor TOP10vector. (Bakker & Kolk 2003)

3.1.1 Characteristics

Top10NL has a number of distinguishing characteristics. The main characteristics, seamless data, object orientation, geometric multiplicity, attribution, planar partition and relative height level, are explained in this paragraph.

In contrast to Top10vector, TOP10NL is **seamless** data. Geographical entities are stored as single objects, no longer split by map edges. (Kadaster 2013, pp.8–9)

Top10NL is an **object-oriented** database. “In an object-oriented description the world is seen as a collection of objects and characteristics which are incorporated in the data-model as entities and attributes… An object is a terrain element that has an identity, about which we can talk and that can be manipulated.” (Bakker & Kolk 2003, p.843) However, the definition of an object is ultimately dependent upon the intended application. Bakker and Kolk illustrate this by the example of a bridge. A bridge could be discriminated from a road as an individual object, but equally well as characteristic of that road. It all depends on the ultimate application.

Internal demands at Kadaster as an NMA (production of topographic datasets and maps) require the capturing of geographical objects for the 10K map. These objects are stored in distinguished object classes, see Table 3.

Table 3: object classes in Top10NL

Entities
Road segment
Railroad segment
Water segment
Terrain
Building
Specific terrain element/ construction
Administrative area
Geographic area
Functional area

(Bakker & Kolk 2003; Kadaster 2013, p.18)

Several characteristic of entities are stored as **attributes,** which can be identifying, describing, geometric, meta and temporal. Each object or entity has a unique identifier (UID): TOP10ID, which is unique for the whole dataset, but is meaningless regarding object characteristics. Objects are given meaning by descriptive characteristics, which constitute of classifications such as water types (stored in domains) and unique attributes such as names (Bakker & Kolk 2003). Some attributes can store multiple attribute values.

The geometries of entities from all object classes in Top10NL are stored as points, lines or polygons to define their location and shape. Most objects in the object class Road segments can have **multiple geometries**. Connections can be visualised as area or (centre-) line features, while crossings can be visualised as area or point features (Bakker et al. 2005). Intentions were to adopt the concept of geometric multiplicity for Water segments (Bakker & Kolk 2003), but this did not make it in the eventual implementation.

The object classes water polygons, land cover and road polygons (*waterdeelVlak*, *terreinVlak* and *wegdeelVlak*) together form ***planar land cover***, (Kadaster 2013, p.10) which means the whole of Dutch territory (on the European continent) is covered by a polygon from one of these feature classes, which are topologically related. Besides these topological partition, there are object classes stored separately, and these can be superimposed on the planar partition. Examples are buildings, specific terrain element and administrative areas. (Bakker & Kolk 2003)

Top10NL was initially designed to store Z-values of objects (Bakker & Kolk 2003, p.846), but there has never been an implementation. Instead, ***relative object heights*** are stored in a height-level attribute. This attribute indicates the relative height-level of objects for local situations. Between design and implementation there was a shift in definition values which should be attributed to indicate height-level. By design ground level was intended to be attributed height-level '0' (Bakker & Kolk 2003), but in implementation (almost) all objects visual from air have to be attributed height-level '0' (Kadaster 2013, pp.26–27). This shift is justified by the fact data acquisition is from aerial photography which is not suited for proper identification of ground-level. However, there are developments utilizing laser scanning and point clouds, which culminated in the release of a new product 3DKaartNL, building on the researches of Stoter (2012) and Oude Elberink (2010).

For hydrography height-level indicates locations of bridges and culverts. They should be considered in establishing a hydro-network.

3.2 TOP50raster: a concise overview

The target dataset for the generalisation workflow is TOP50raster, which is one of the three digital map series in the product family TOPraster. TOP50raster is a generalised digital cartographical map, scale 1:50.000. It is produced for use as overview map in the Military, but also as reference data in GIS systems. The product is derived from a cartographic object oriented database, which inherits the characteristics from TOP10NL. This dataset is not suited for GIS-analysis due to applied generalisation techniques like displacement and omission. The focus of the product is to provide an intelligible overview of an area.

3.3 Preceding work on automatic generalisation in the Dutch context

In 2013 interactive generalisation was replaced by an automated procedure, after an extended period of research and development. The prior research developed in five phases.

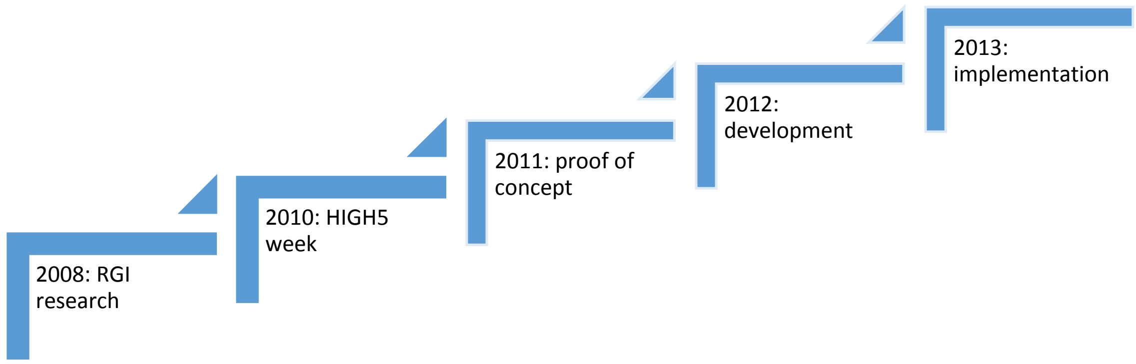


Figure 26: phases in the research and implementation of automatic generalisation at Kadaster

Within the project ‘Ruimte voor Geo-Informatie’ a group of researchers (from industry and NMA) investigated automatic generalisation of TOP10NL to derive TOP50NL data. (Hardy et al. 2008) The objective of this study was to answer the question “what automatic generalisation could offer, taking the current object models, products and generalisation-specifications for granted”. A number of experiments were set up, which focussed on roads and buildings. Water network thinning was neglected. (Stoter & Smaalen 2008) Results of these experiments were very promising. A consequent study took previous results and experiences and investigated how both can inform Map requirements and specifications to fit automated generalisation. (Stoter et al. 2009).

The next research phase was in 2010 when a team of experts performed an innovation sprint within five weeks. This innovation concept is called a HIGH5. The objective of the HIGH5 was to generalise a 50k visualisation fully automatically from TOP10NL data with available commercial tools within a set amount of time. The results of this feasibility study should provide insight in the potential of fully automatic generalisation. (Bruns et al. 2010, p.3). It was explicitly stated there was no intention to copy the existing 50k map, but the main generalisation principles should be acknowledged. (Bruns et al. 2010, pp.3–4) In this phase the first attempts at Kadaster were made to automatically generalise waterways from 10k to 50k. (Bruns et al. 2010).

In the subsequent phases, (proof of concept and development), the workflow was expanded and became more mature: it covered a greater diversity of generalisation issues and the generalisation workflow was also tested on other areas in the Netherlands. By implementation of multicore processing the workflow became mature enough to replace the existing interactive generalisation workflow.

3.4 Resulting workflow

Starting point for this research is the current implementation of generalisation at Kadaster : a TOP10NL dataset which has been prepared by data validation and data-enrichment and has been translated to the data scheme of TOP50 (Stoter et al. 2011). This is where the generalisation starts. The generalisation process can be divided in three sub steps (Figure 27): geometric generalisation, displacement and visual enhancement. The thinning of the water network is embedded in the first sub step, geometric generalisation.

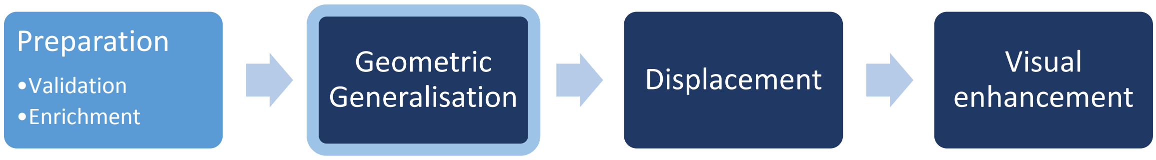


Figure 27: overview of generalisation process

The procedure for waterline features in the geometric generalisation is indicated by the workflow in Figure 28, which can be divided in five subsequent actions.

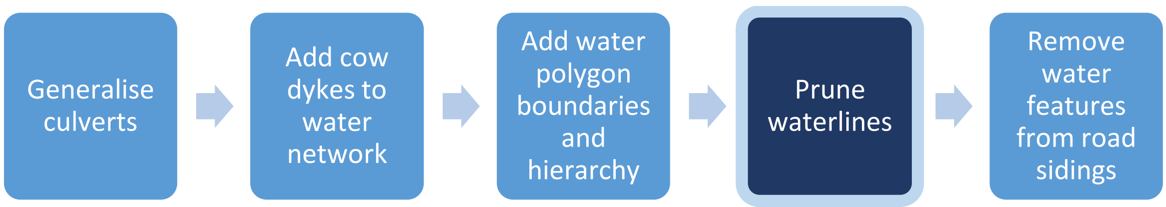


Figure 28: current implementation of water thinning

1. Generalisation of culverts. In the traditional generalisation-specification there are rules which require the translation of culverts to normal waterways under specific constraints.
2. Add cow dykes features to complete the water network. *(dam, koedam*) Cow dykes are small artificial connecting lines to the water features.
3. Add the boundaries of water polygons as lines to complete the water network and populate a hierarchy field for all water features to ensure a certain hierarchy.
4. Prune the water network by application of Esri's Thin Road Network-algorithm with a minimum length parameter value of 300 meter.
5. Erase water line features overlapping with road sidings since these features because graphic conflicts on the map.

The experiments in the next chapter extent the second and third action and research alternating thinning algorithms (action 4). The last action is out of scope in this research.

3.4.1 Partitioning

One additional concept is important to discuss before advancing to experimentation: partitioning. This addresses the impossibility to process the entire dataset for the Netherlands at once. Partitions divide the source data into smaller subunits and are constructed from artificial and existing topographical (linear) objects which are important enough to be maintained during the process and delineate a reasonable area and number of features.

The set of partitions should cover all features in the dataset. The inner boundaries of the partition dataset are constructed from the main road network pattern, the outer boundaries from the nation's border or the coastline. Some artificial connections between the inner partition boundaries and the outer boundaries should be added (Stoter, Post, et al. 2014), see Figure 29.

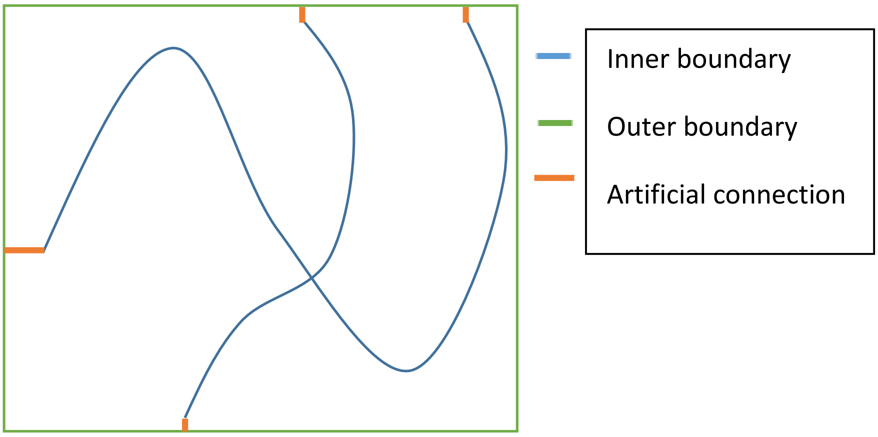


Figure 29: extension of line elements to partition borders

The process results in approximately 480 partitions which can be processed individually.

Figure 30: the Netherlands divided in partition for processing

4 Experiments, results and findings

This chapter is structured in five sections. The first section (4.1) explains the prototyping workflow. In the next section the analysis of landscape characteristics will be addressed in several experiments. The main quest in these experiments is to establish the usability of the landscape typology dataset of CHA-NL in generalisation experiments (4.2) and to select representative datasets for experimenting (4.3). The third section (4.4) discusses the implementation of evaluation algorithms which will be used in the subsequent experiments to evaluate the results of generalisation approaches. This finishes the preparatory experiments.

In the following section an algorithm is researched to improve network connectivity (4.5) and three different generalisation algorithms are applied (0). Both experiments are applied to the selected test partitions to get results for different landscapes.

The final section contains the evaluation of findings (4.7) for each generalisation algorithm and an equation of results between landscape types and applied generalisation methods.

4.1 Prototyping workflow

Each aforementioned experiment addresses a part of the research question and together they form the prototyping framework. The coherence of experiments in this framework is explained in Figure 31. In the framework three components can be distinguished: preparation, thinning and evaluation.

The Preparation algorithms are used to investigate the initial situation of the test data, to enrich water features with landscape type, to enhance network connectivity of the water features and to identify test areas. In subsequent processing this information will be used as input and benchmark data. Part of the preparation is an evaluation of network improvement using three metrics: (1) an equation of the number of networks in the initial and improved situation, (2) the identification of the connectivity level of networks and (3) the computation of the number of circuits within a network. Metric 1 is used to evaluate the improvement of the preparation; metrics 2 and 3 are used for comparison in the subsequent stages. Last two metrics are identified as network statistics

The Thinning component applies three different pruning algorithms to the test areas. The third and last component evaluates the results of the three pruning algorithms on basis of network statistics for each individual landscape type, but also compares the results against each other and the base dataset.

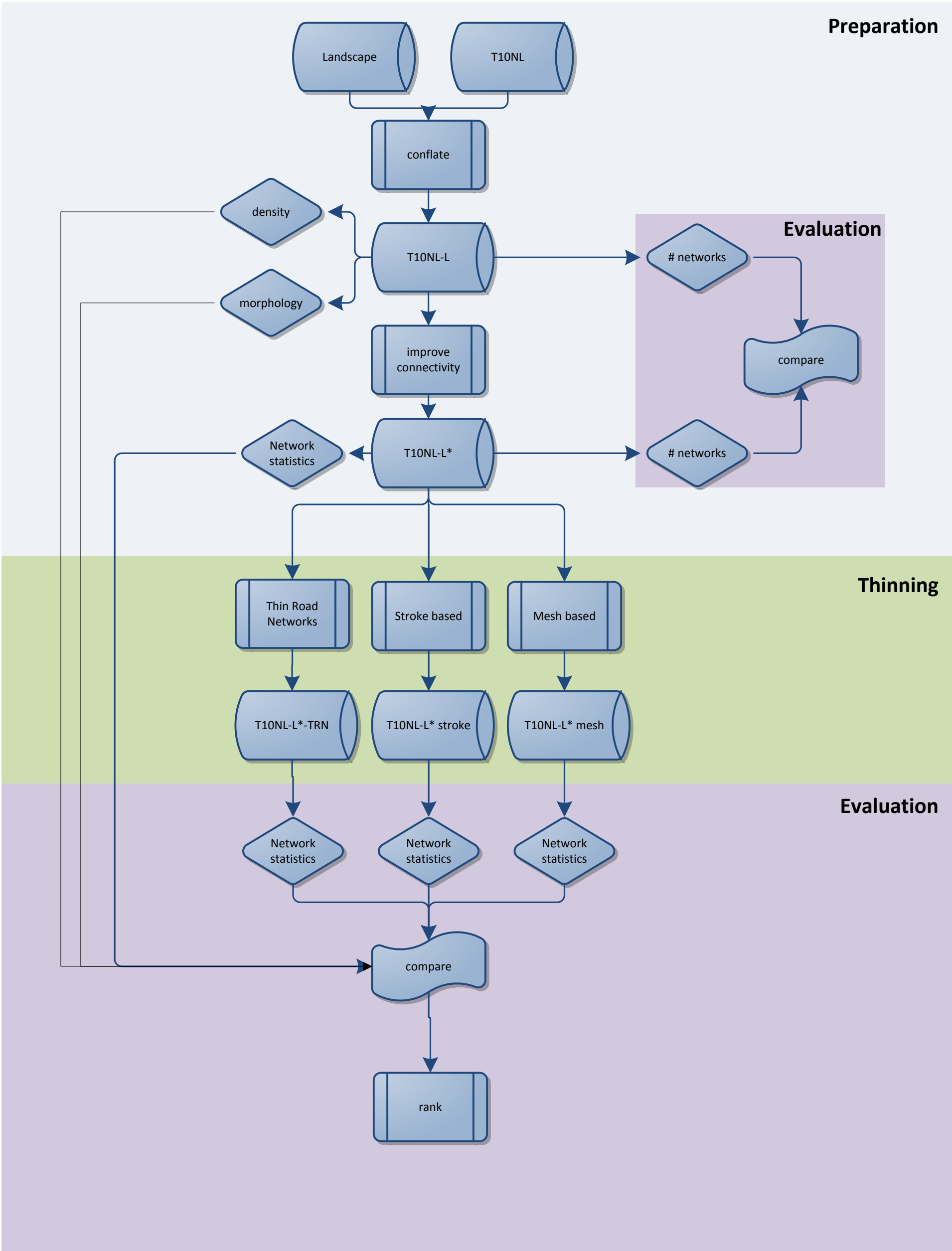


Figure 31: prototyping workflow

All algorithms (with one exception which contains a Python script) are created using ArcGIS Modelbuilder. ArcGIS models use the following symbology convention:

Figure 32: ModelBuilder symbology convention

Instead of providing all the technical details, this chapter contains a summary of the algorithms in conceptual images. This aims at avoiding reader confusion with an overload of technical details. The symbology convention used in the conceptual images is displayed in Figure 33. Intermediate output which serves as input to subsequent processes will be neglected in these images.

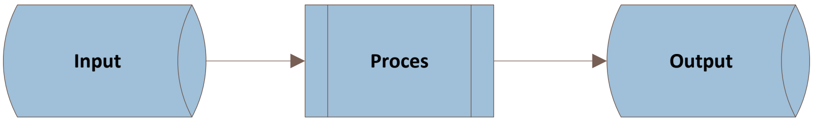


Figure 33: conceptual images symbology convention

Both models and script are included in full detail in the appendices.

4.2 Analysis of landscape characteristics

The attribution of landscape type to water features is based on an existing classification. The landscape type dataset as provided by CHA-NL is evaluated by two criteria to determine if this dataset is a usable source to define landscape type of a partition. The first criterion is to calculate the humidity factor for each landscape type. The second criterion is to calculate a feature morphology index. The paragraphs below provide a description of both methods.

4.2.1 Calculate humidity factor for each landscape type

The humidity factor is the ratio between the area covered by water features and the overall area of the landscape type and ranges from 0 to 100. It can be used to indicate the overall density of water for a specific area. Landscapes which are not covered by water features have a water density of 0, while a density of 100 indicates a landscape completely covered by water features.

These figures can be used to characterise the water areas. An approximation of area A covered by water features which are represented as waterlines in TOP10NL (waterlines with a width smaller than 6 meter) is calculated using the formula

$$A = l * w$$

where l represents the accumulated length of all water features and w represents the assumed average width of the class. Dependent on class type the value of w varies according to the values in the table below.

Table 4: assumed average class width

class type	**assumed average width (w)**
greppel	1.5
sloot, 0,5 – 3 m	
sloot, 3 – 6m	2.5

The statistics for wet areas are calculated using the formula:

$$wet\ areas = \frac{Agreppel + Asloot0{,}5 - 3m + Asloot3 - 6m + AwaterAreas}{totalArea}$$

4.2.1.1 Algorithm

The model *calculateWaterStatisticsForLandscapeType* (included in Appendix A) calculates both the percentage of the area covered by water areas, small waterlines, wide waterlines and trenches and the ratio of wet and dry areas for each landscape type.

The algorithm can be broken down in three subsequent actions. Water features and Landscape-type features are conflated using the intersect tool of ArcGIS. This information is used to calculate the surface of the areas covered by water and in this action information is retained about the original water features. The last action in the algorithm makes all the information relative by calculating percentages which can be used in equations.

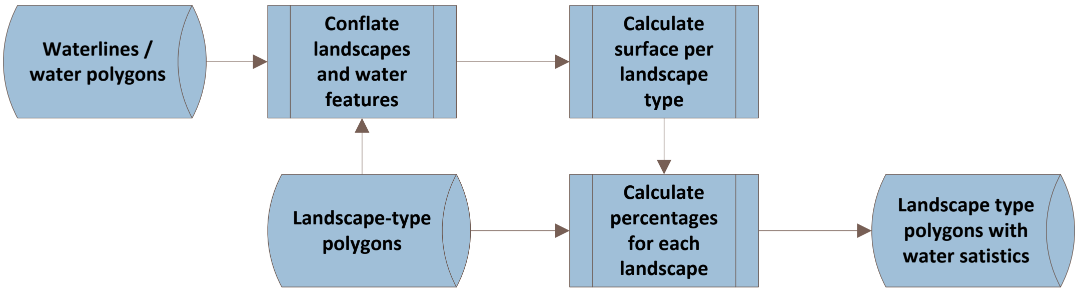


Figure 34: conceptual overview of the calculateWaterStatisticsForLandscapeType-algorithm

4.2.1.2 Results

The algorithm results in statistics on humidity of landscape – type areas and in statistics which reflect the composition of water features per landscape type. The results of the algorithm are portrayed in the graph in Figure 35. This shows the distribution of water type classes for each landscape type.

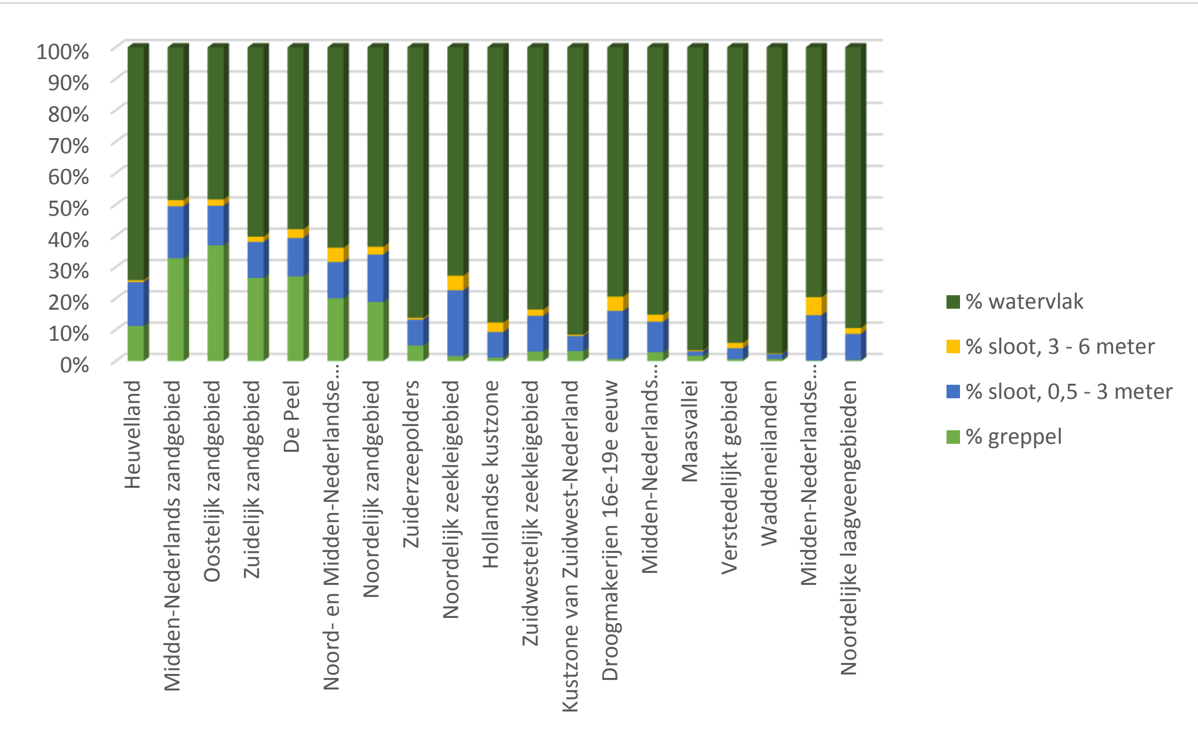


Figure 35: composition of wet feature types per landscape type

The humidity level is portrayed as a choroplet map in Figure 36.

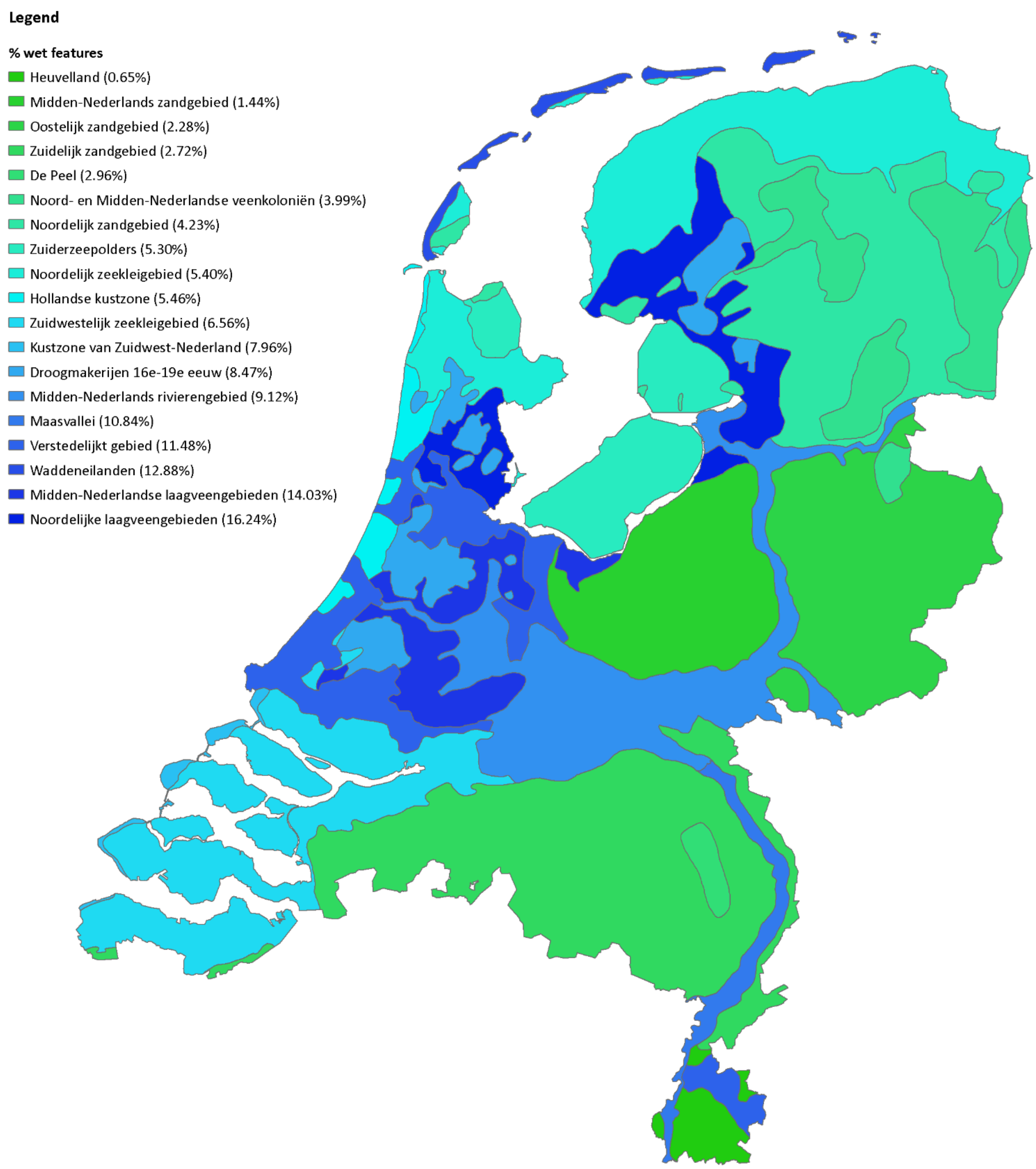


Figure 36: humidity per landscape type

These results show that the landscape typology classification of CHA-NL contains the necessary variation in humidity.

4.2.2 Establish geo-morphology of line features

Besides humidity, feature morphology also provides information about landscape type. The main idea here is that some landscape types will have mostly straight lines, e.g. polder landscape or former histosols (areas with moor), where other landscape types will have more naturally formed winding lines, e.g. hilly areas. It is possible to calculate a feature morphology index per landscape type. The feature morphology index M is the ratio between straight and winding features within a given area and can be expressed as

$$M = \frac{f}{v} * 2$$

where f is the total number of line features within the area and v is the total number of vertices for these features. Because a straight line always has a starting and an end point, a multiplication by 2 is performed. An M-value close to 0 indicates an area with almost only straight features, whereas areas with an M-value close to 1 consist mainly out of winding features.

4.2.2.1 Algorithm

This algorithm computes a feature morphology index for a given set of line features and can be found in Appendix B. The concept of the feature morphology index is explained in 4.2.2. The technical implementation is explained here. Prerequisite to this tool is the conflation of landscape type polygons with waterline features (see 4.2.1.1).

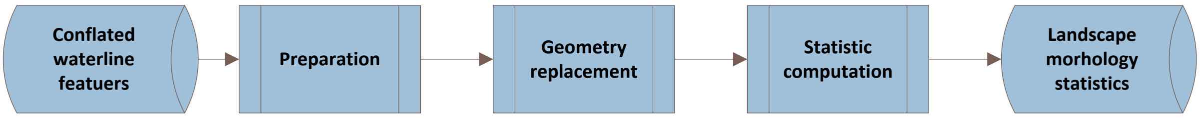


Figure 37: conceptual overview of feature morphology algorithm

The tool iterates the different landscape types selecting all waterlines belonging to one landscape type. For each iteration it performs the following actions for the selected landscape:

- The input data is prepared by dissolving all connecting elements in the same class (not permitting multiparts) to create longest possible lines and remove pseudo nodes. (For the Dutch situation this classes are created from the unique combinations of the attribute values for water type *(type water)* and width class *(breedteklasse)*. The second preparation action is the removal of bias caused by unnecessary detail in the line geometries. To achieve this, point reduction is applied (using the Douglas Peucker algorithm as applied in the Simplify Line tool of ArcGIS with a threshold value of 2 meters). The results are only used in this algorithm.
- All resulting line features are fed into the Feature Vertices To Line-tool, which creates a point feature for every vertex. For each point feature the parent attributes are stored, which can be used to diversify statistics.
- For the given set, statistics are calculated for the total number of features, the total number of vertices. These statistics are also computed per class (using the classification indicated above). The final step is to calculate a feature morphology index for each class and for the total line set.

4.2.2.2 Results

The results of the algorithm are displayed in Figure 38.

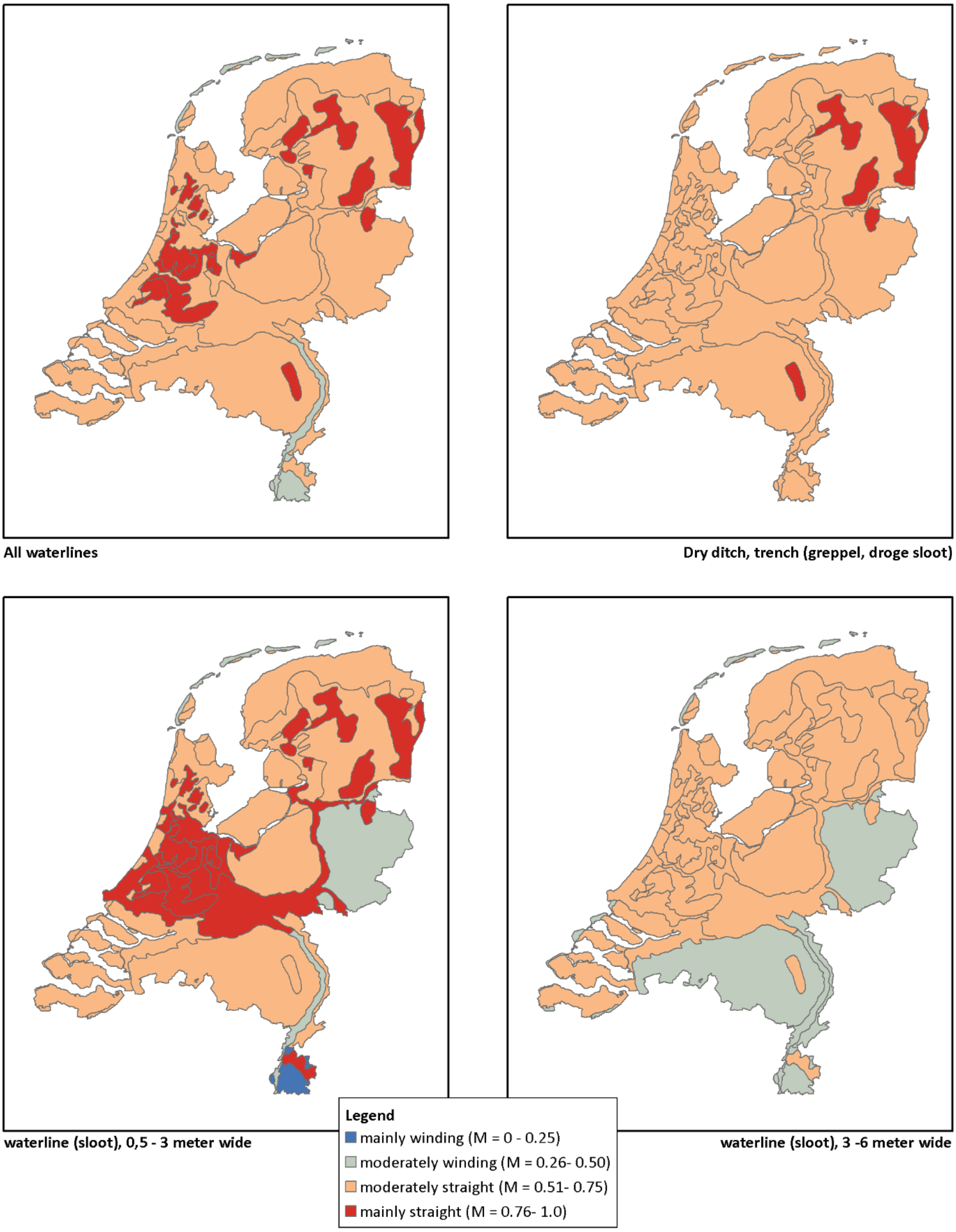


Figure 38: feature morphology index per landscape area for all water features and diversified by water type

4.2.3 Conclusion

The results of the two preceding experiments show the landscape typology dataset of CHA-NL is a usable source to define landscape type of a partition. The typology offers enough variation in humidity-level and average feature morphology to be used as context-identifier.

4.3 Selection of test partitions

As discussed in 3.4.1, the source dataset TOP10NL is divided in approximate 480 partitions for processing purposes. By performing a spatial overlay with landscape type features, the partitions can be classified according to landscape. Although the spatial relationship between partition and landscape type can be M: N, there are partitions which are fully contained by one landscape type polygon. For each landscape type the smallest partition from the subset of single-landscape partitions is selected as candidate for prototyping.

4.3.1 Algorithm

The algorithm (full details in Appendix C) intersects landscape areas with partitions, which results in new polygons. The number of landscape types for each original partition is established and each original partition with only one landscape is a candidate for prototyping. The last action in isolating test partitions is the selection of the smallest test area per landscape type.

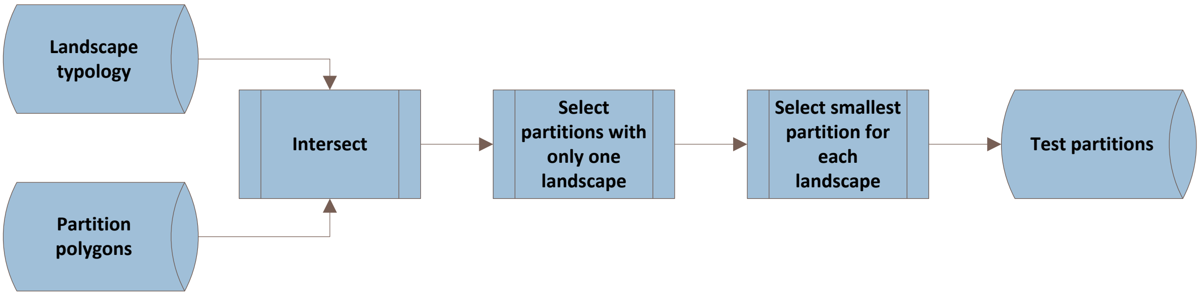


Figure 39: conceptual model of algorithm to select test partitions

4.3.2 Results

The algorithm results in the selection of twelve partitions with a unique landscape type (Figure 40). Seven unique landscape types are omitted (Table 5) as they do not contain complete partitions. These landscape types are excluded from prototyping. These are not necessary to prove the concept of landscape-typology as context identifier.

Table 5: omitted landscape types

De Peel
Droogmakerijen 16e-19e eeuw
Hollandse kustzone
Kustzone van Zuidwest-Nederland
Midden-Nederlandse laagveengebieden
Noordelijke laagveengebieden
Waddeneilanden

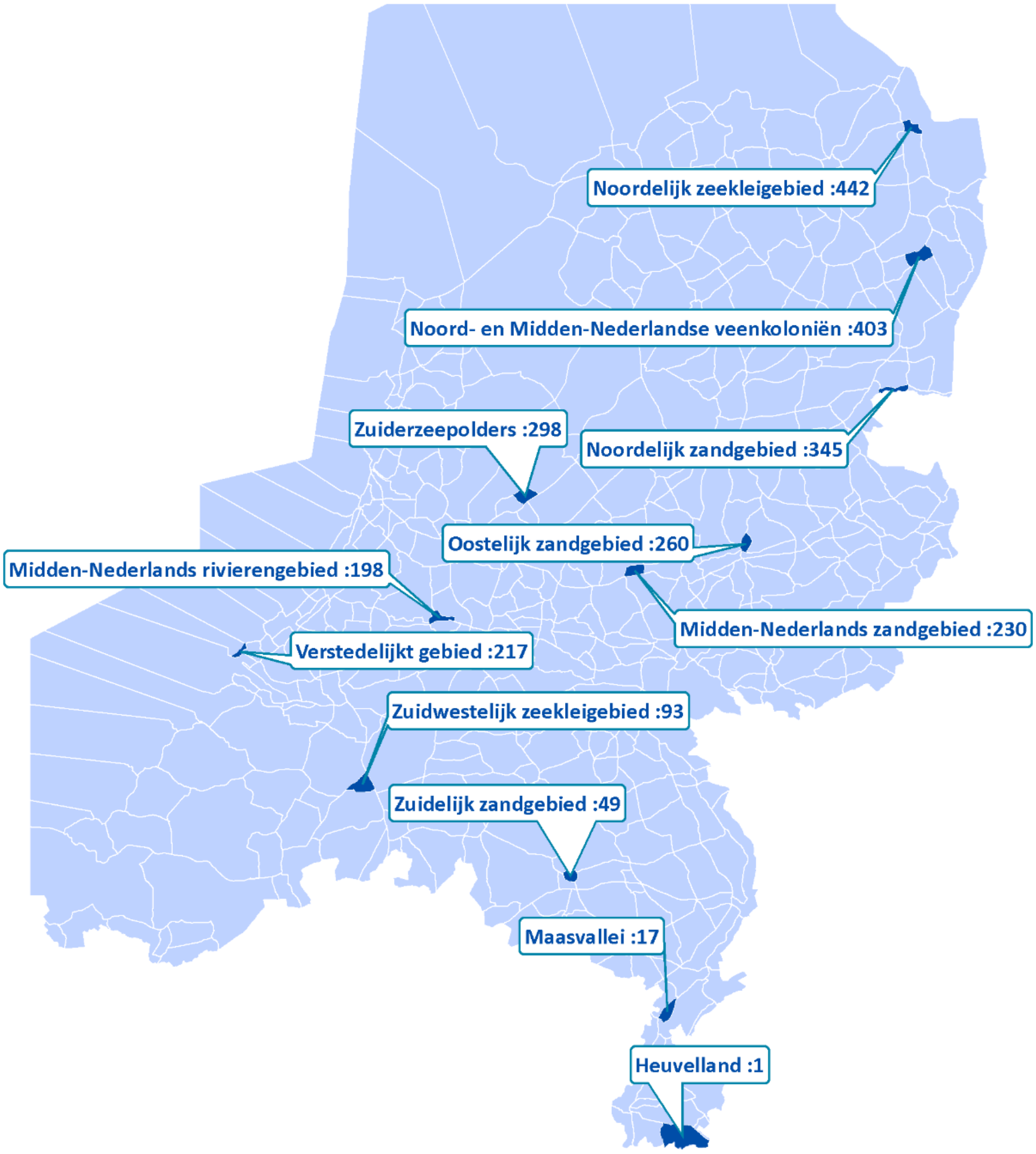


Figure 40: preliminary test partitions

4.4 Evaluation algorithms

The prototyping framework described in section 4.1 also includes several evaluations of intermediate results and a comparison between intermediate and final results. Three metrics are proposed and the corresponding algorithms are explained here. Discussions of results are included at later stages where the algorithms are applied.

4.4.1 Count number of networks

The evaluation algorithm *numberOfNetworks* (Appendix D) computes the number of networks within the dataset. This is used to evaluate the improvement of the network connectivity. The concept is fairly simple: buffer all features with a distance of 1 meter (an arbitrary value) and allow for multipart features. This will result in one multipart polygon for all present water features. This multipart polygon is subsequently broken into multiple features. Since this new features are not connected, they form a good representation of the number of networks.

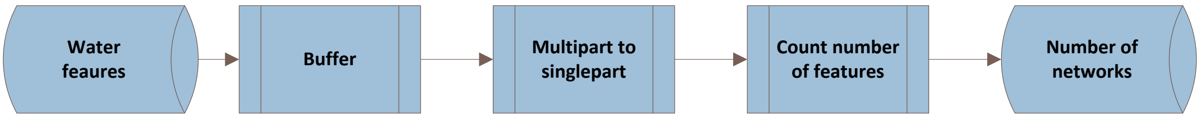


Figure 41: number of networks algorithm

The focus of this algorithm is on a dataset. It informs about the number of networks, but it does not provide details about connectivity or structure of each network.

4.4.2 Network connectivity statistics

Network statistics focus on the connectivity level (γ - index) and the number of circuits (α- index) for a network and are an implementation of the theory as described by Wong (2005), see also section 2.6.1.2 on Network Statistics. The algorithm (Appendix E) can be summarized at conceptual level as shown in Figure 42 below. Prerequisite to the algorithm is a planar topological graph.

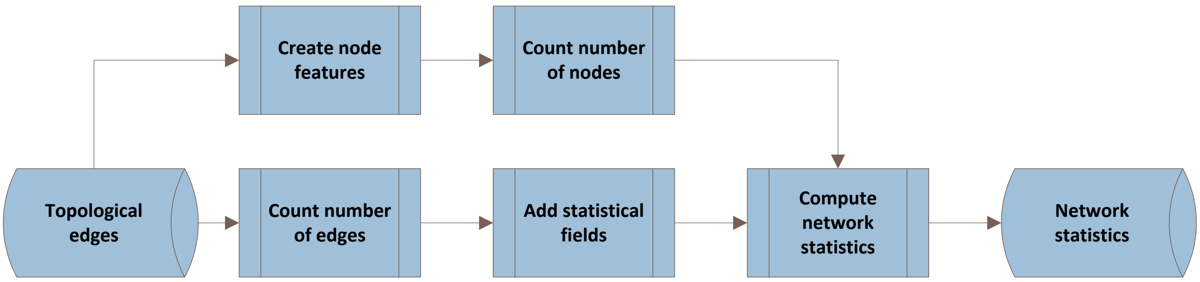


Figure 42: network statistics algorithm

These metrics will be used in the following experiments on base-data-improvement and in the evaluation of thinning algorithms.

4.5 Base data enhancement: improve connectivity

One of the assumptions in this thesis is that better connected networks will deliver better results in thinning. It is appropriate to elaborate a little on the concept of “better connected networks”. The underlying idea is that in the human mind choices for thinning are not solely determined by the existence of connections in the terrain, but also (and maybe foremost) by the perception of continuous structures on a map. Interrupted lines (for example, watercourses in extension of each other, but interrupted by a road) may form a closed network in the human perception, although no physical connection exists, as illustrated in Figure 43.

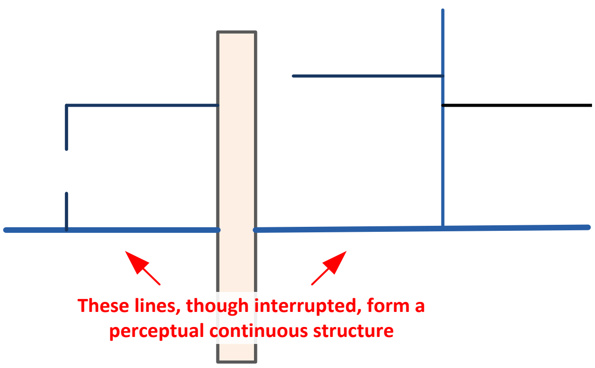


Figure 43: a cartographer perceives a continuous network, though the physical connection is missing

The following paragraphs describe an experiment which focusses on the issue of better-connected-networks and tries to simulate the human's perception. Figure 44 below shows a conceptual representation of a fictive topographical situation. It contains ditches, small and broad waterlines, isolated and linked water areas, and a road.

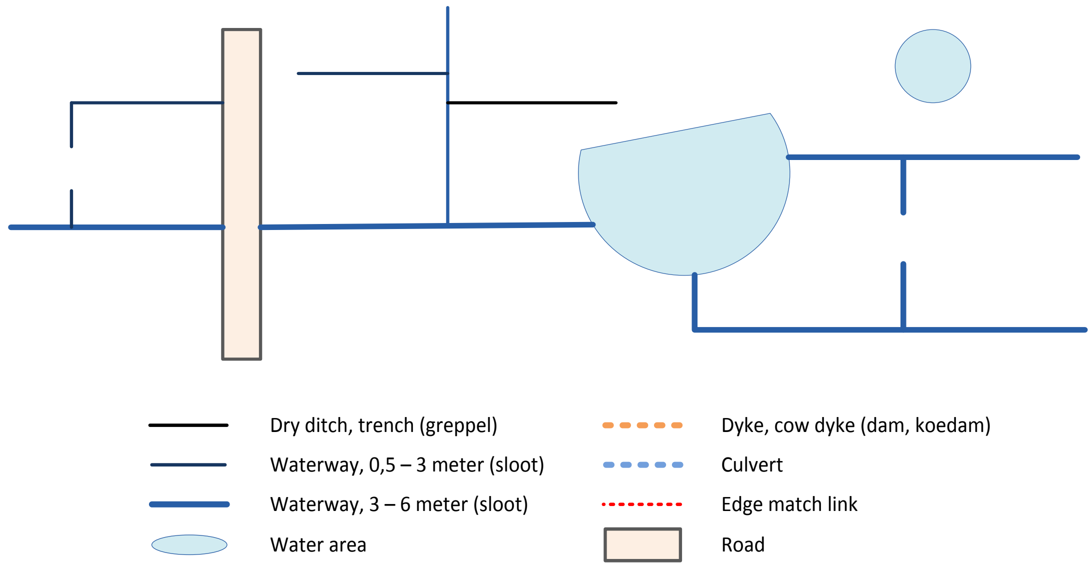


Figure 44: Initial topographical situation and Legend for Figure 44 to Figure 50

From this initial situation networks can be created, but parts which belong to the perceptual network remain unlinked, as shown below in Figure 45.

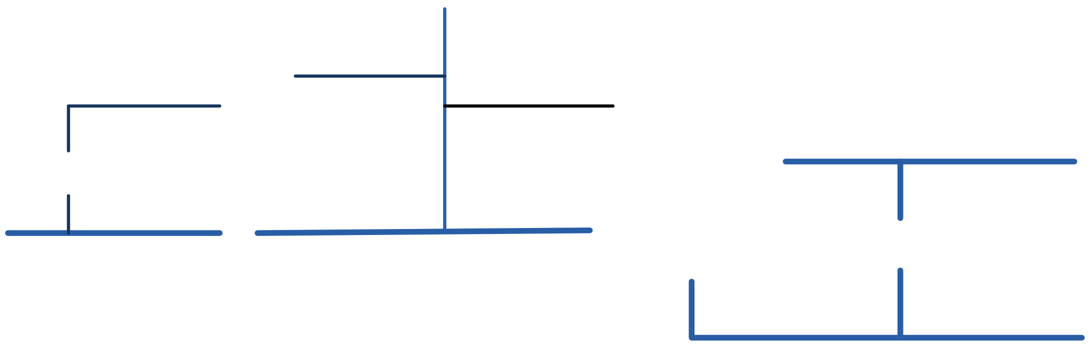


Figure 45: initial network form waterline elements

4.5.1 Improving connectivity of a water network from TOP10NL

There are two approaches to improve the connectivity of the network. By adding (geometrical) information which is present in the geographical dataset (Heinzle et al. 2005) and by constructing perceptual links. The addition of culverts and cow dykes to the network belong to the first approach; the calculation and addition of artificial connecting lines (replacing water areas by centrelines) and perceptual links (see below for an explanation) belong to the second one. The current Kadaster workflow has already implemented the addition of culverts and cow dykes.

4.5.1.1 Addition of culverts

Culverts (*duikers*) are tubes to connect watercourses below the earth surface. Their main function is to maintain the connection between waters or drainage of adjacent lands, cf. Winkel (2012, pp.15–

5): *"…de instandhouding van de verbinding tussen wederzijds gelegen wateren of de afwatering van aangrenzende landerijen."* Although culverts are invisible on aerial photographs, they should be used in water network construction. This is acknowledged in the design of the Top10NL data model, which includes culverts in the watercourse lines. Figure 46 illustrates the network becomes less fragmented after addition of culverts. The number of networks is reduced from 5 to 4.

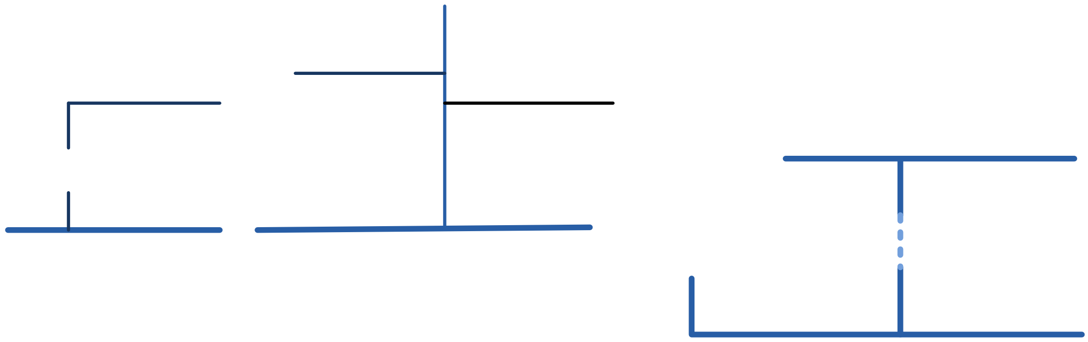


Figure 46: add culverts to network

4.5.1.2 Addition of cow dykes

Top10NL also includes the object cow dyke (*dam*, *koedam*) which contributes to the construction of a water network. Cow dykes are earth bridges which connect two pastures (or a pasture and a road) across a stream or ditch, to allow cows to move between pastures. They often contain a culvert pipe to allow water to pass underneath, cf. Winkel (2012, pp.15–1): "*Ontsluiting van een perceel, dwars over een waterloop, door middel van demping (al dan niet met doorlaatbuis)*".

Surprisingly, given their consistency with water features, cow dykes are not included in Top10NL in the watercourse lines, but in the object class specific terrain element/ construction. To construct the water network this objects should be added to the watercourse line object class. The addition of cow dykes again defragments the network, reducing the number of parts from 4 to 3, see Figure 47.

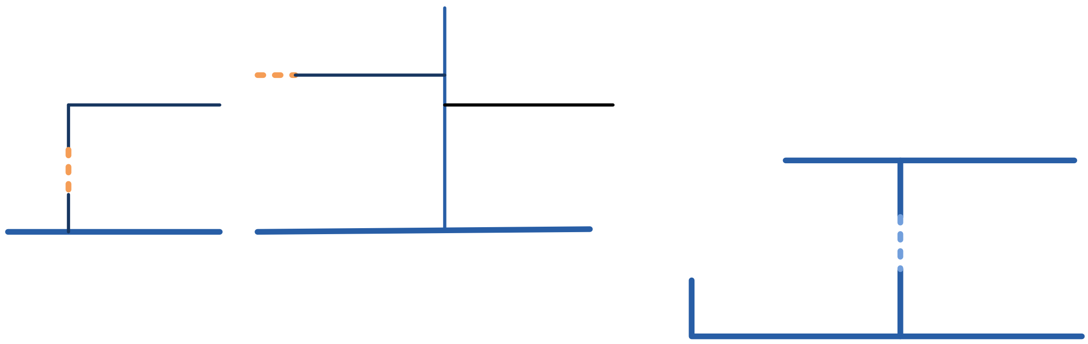


Figure 47: add cow dykes

4.5.1.3 Replace water areas in the network with connecting lines

Water areas are retained in the resulting 50K map. However, because they are important parts of the water network they should be used in the construction of the network. Because the tools cannot handle features with different geometry, water polygons have to be translated to polylines. In this research polygons are replaced by their calculated centrelines while retaining the topological relationship with adjacent waterlines. Results of this approach showed to be most suitable (they did not cause crashing algorithms) and efficient (they resulted in clean geometries) in pruning algorithms. Figure 48 illustrates the reduction of networks from 3 to 2 parts, after addition of water polygon centrelines to the network.

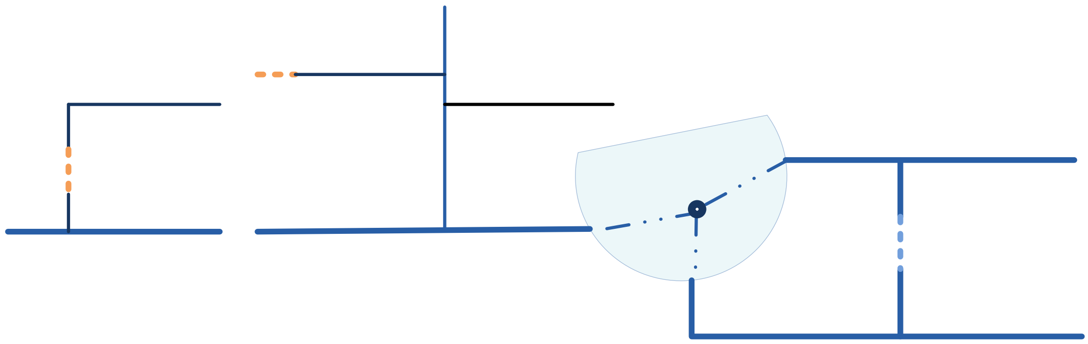


Figure 48: replace water areas in the network with connecting lines

4.5.1.4 Calculation of perceptual links

The connectivity of the network can be further improved by the introduction of perceptual links, which has already been advocated above. To summarize this concept: two waterlines in extension of each other, but divided by a road are perceived by the human eye as a continuing line in a network. By creating lines in the database for these virtual elements, the connectivity of the network can be improved. In Figure 49 the addition of a perceptual link reduces the number of networks to 1, providing better connectivity for waterlines.

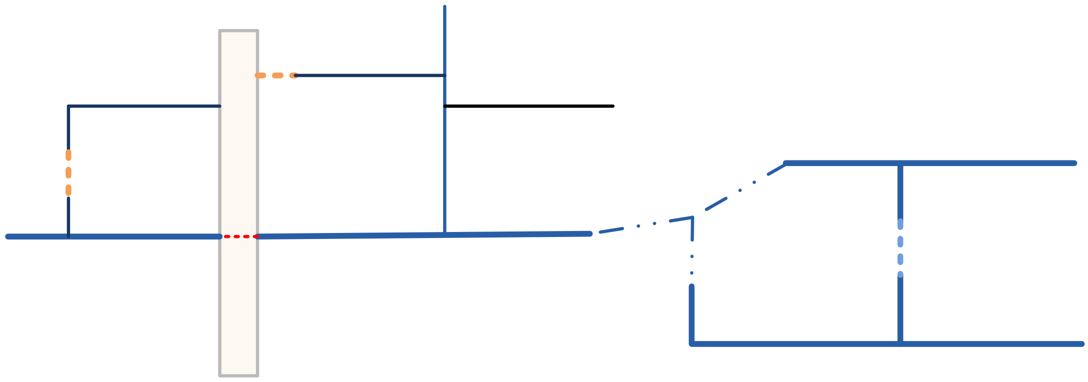


Figure 49: add perceptual links

Using these preparatory steps improves the connectivity of an initially fragmented water network, as the final result in Figure 50 shows.

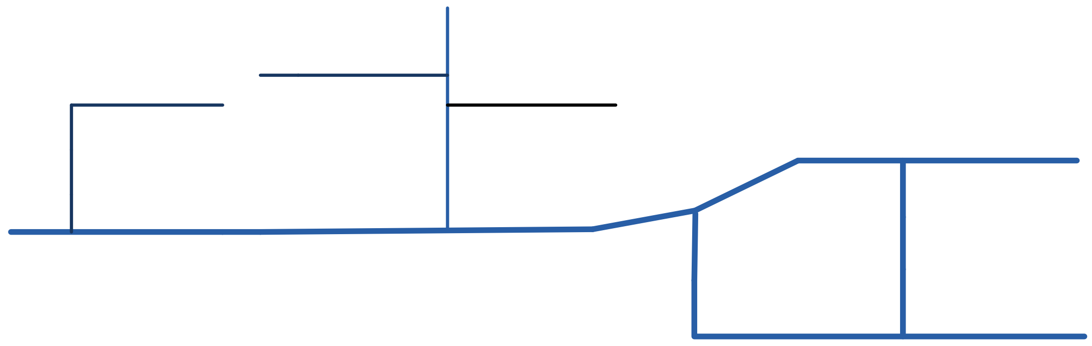


Figure 50: resulting network

4.5.2 Algorithm

The concepts explained in paragraph 4.5.1 are implemented in an algorithm which is explained in Figure 51 (for full implementation, see Appendix F). The algorithm starts with data directly taken from Top10NL: water areas (*waterdeelVlak*), waterlines (*waterdeelLijn*) and a subclass of specific terrain constructions (*inrichtingselementLijn*). This data is topologically checked, connected water areas are replaced by their centreline, all line features are merged into one dataset and the algorithm concludes with the construction of perceptual links to create an improved network.

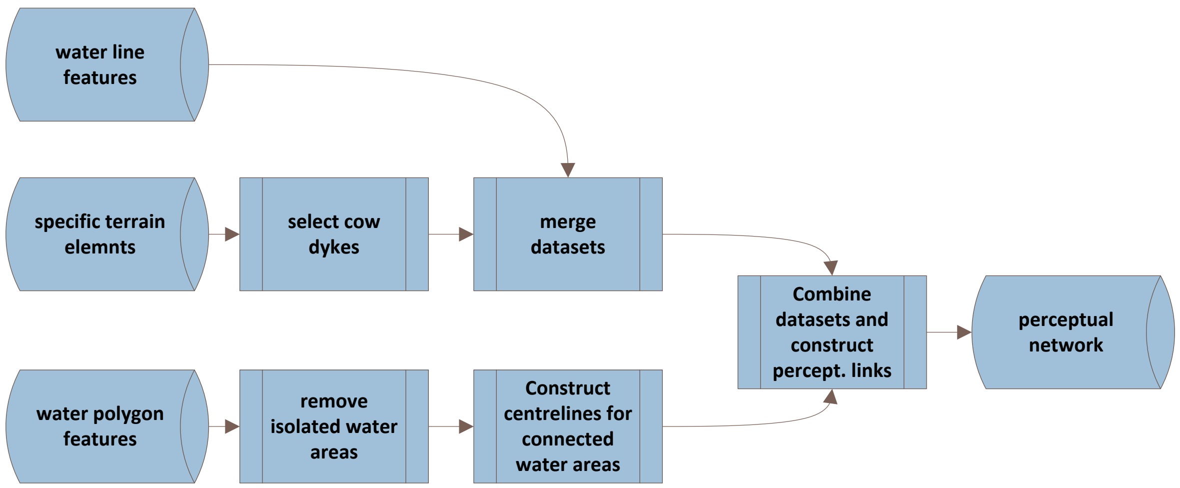


Figure 51: algorithm to improve network connectivity

4.5.3 Results of the connectivity improvement algorithm

The algorithm was executed on all partitions and the results are evaluated using the countWaternetworks tool. Table 6 shows the results of all subsequent processing actions and provides information on intermediate and final results of the algorithm. COUNT gives the number of networks for the test data. Ideally, this number should reduce after each processing action. MIN displays information about the number of features within the smallest network; MAX indicates the number of connected features in the largest network (and during processing this number should increase); the final metric MEAN provides statistics on the average number of features for all networks in the specific area.

Table 6: results of improving the network

	WaterdeelLijn				WaterdeelLijn + Koedam				WaterdeelLijn + Koedam + WaterAreaCenterLine				WaterdeelLijn + Koedam + WaterAreaCenterLine + Artificial RoadConnections			
partitie	**COUNT**	**MIN**	**MAX**	**MEAN**	**COUNT**	**MIN**	**MAX**	**MEAN**	**COUNT**	**MIN**	**MAX**	**MEAN**	**COUNT**	**MIN**	**MAX**	**MEAN**
1	171	1	123	6	159	1	126	6	120	1	487	9	118	1	493	10
17	76	1	156	4	70	1	166	4	69	1	178	4	68	1	178	4
49	19	1	23	3	18	1	23	4	17	1	23	5	15	1	23	6
93	177	1	384	8	99	1	528	15	60	1	1410	27	58	1	1458	28
198	396	1	139	5	168	1	514	14	56	1	2280	51	48	1	2691	61
217	54	1	78	5	51	1	78	5	26	1	162	12	23	1	165	15
260	189	1	103	3	167	1	128	4	120	1	464	7	112	1	513	8
298	219	1	26	2	200	1	26	3	179	1	110	3	148	1	155	5
345	277	1	204	4	179	1	334	6	165	1	334	7	140	1	650	9
403	492	1	74	2	387	1	107	3	267	1	795	6	241	1	971	7
442	255	1	79	4	163	1	249	7	110	1	322	12	99	1	367	14

The results can be displayed in a more transparent way by looking at the reduction of the number of networks for test areas in each subsequent action, see Figure 52.

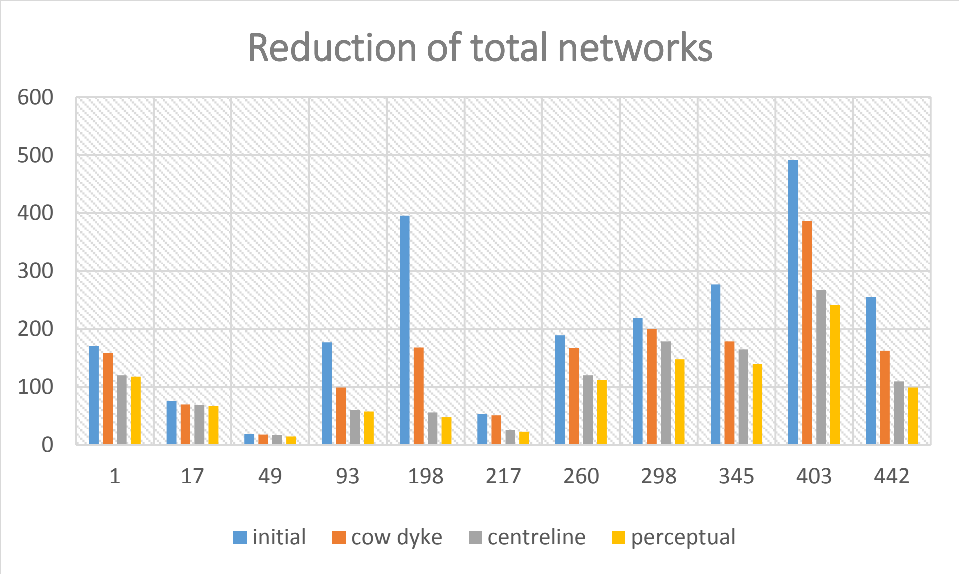


Figure 52: effect of network connectivity improvement on number of networks for each test area

The reduction in network numbers indicates the connectivity of features improved. This indicates the increase of the number of features per network. Both graphs below (Figure 53 and Figure 54) display this development: the number of features for the largest network in each test area increases after each improvement step.

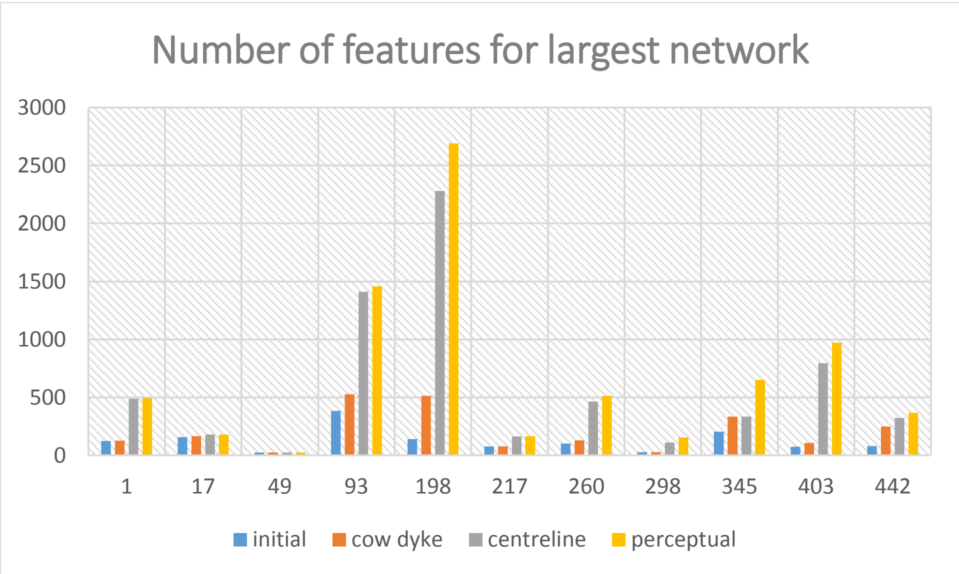


Figure 53: effect of network connectivity improvement on the number of features for the largest network

In general the same is true for the average amount of features for all networks in the test areas. It is however important to bear in mind the possible bias in these results.

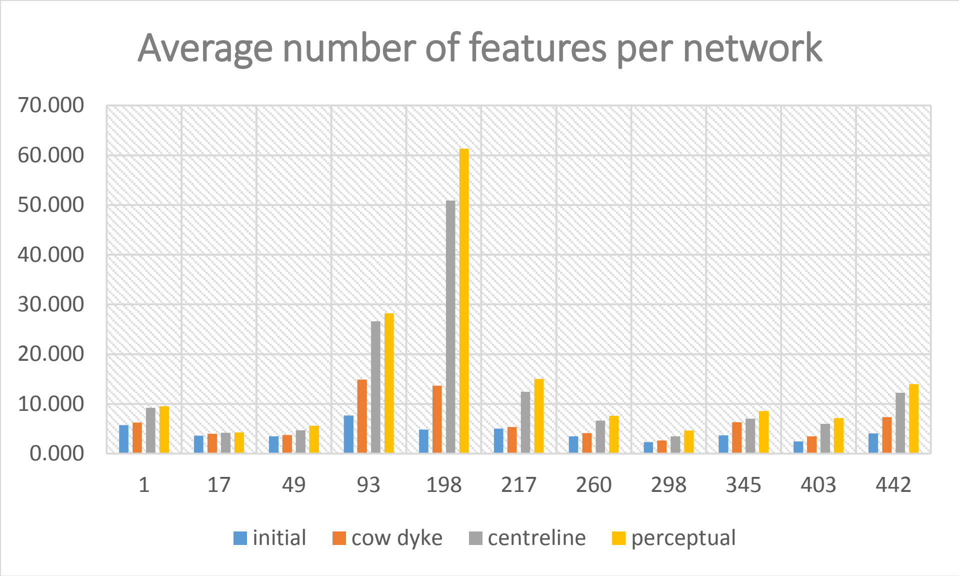


Figure 54: effect of network connectivity improvement on average number of features in networks

In conclusion, the results of this algorithm show significant improvement in network connectivity. However, not all features are connected and isolated features still exist, which is indicated by red circles in Figure 55 for one of the test areas (partition 1). Maps for all test areas are includes as "Input data" in Figure 91 to Figure 101 in Appendix J.

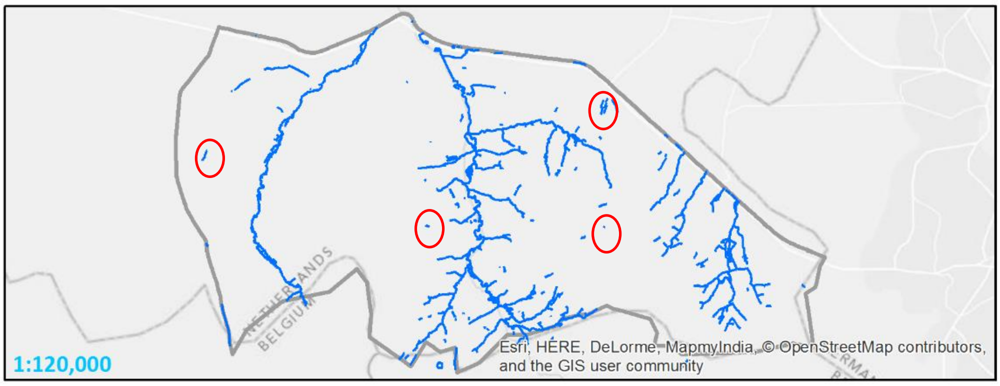


Figure 55: results of improved network algorithm

4.6 Thinning experiments

In the following paragraphs, the results of three thinning algorithms (Thin Road Networks, Stroke-based and Mesh based) are analysed using simple statistical analyses. All thinning experiments start with the results of the improve-connectivity-algorithm (paragraph 4.5 and Appendix F). The results of the experiments are represented as maps in Appendix J and Appendix K and are statistically evaluated using the proposed metrics in the text below

4.6.1 Thin Road Network approach

As discussed in paragraph 2.4.6, the commercial software package ArcGIS offers a network thinning algorithm. Although not exhaustively documented, the results of the tool can be compared with the outcomes of other thinning algorithms and a comparison can be made between the results in test areas with different landscape characteristics.

4.6.1.1 Algorithm

For this research the Thin Road Network-tool of ArcGIS has been used, but it was expanded with a preceding preparatory step (see Appendix G).

This preparatory step can be divided in three actions. First, the input feature class is copied to secure input data from being edited; next, two fields are added for processing purposes to the copied feature class (hierarchy field and invisibility field); and last, all values for the hierarchy field are set to '1', which attributes the same hierarchy value to all features.

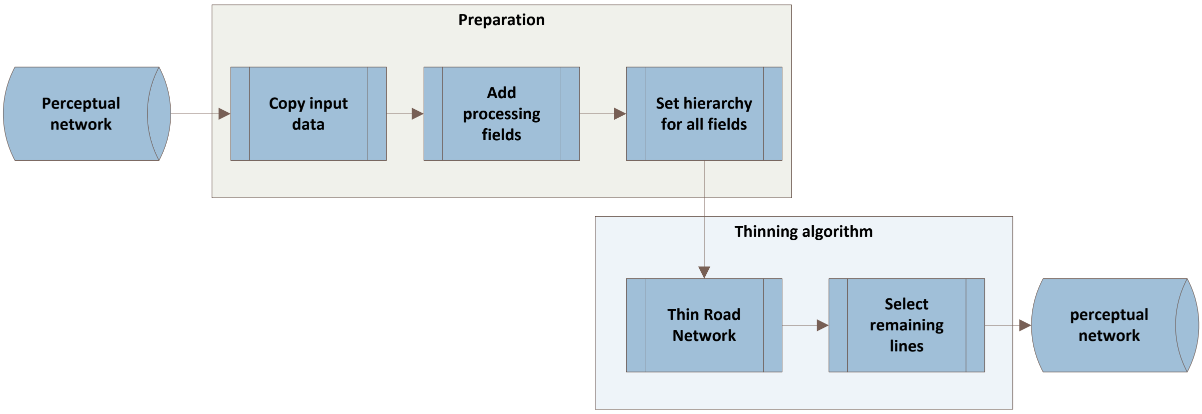


Figure 56: Thin Road Network-based algorithm

The prepared data is fed into the Thin Road Network. The minimum length –parameter for features to retain is set to 300 meters, which is the same value as is used in the current Kadaster implementation process and which is deducted from the manual specifications.

4.6.1.2 Results of Thin Road Network algorithm

The results obtained from the Thin Road Network algorithm are shown for each test partition in Table 7. This table (as well as Table 8 and Table 9 below) has the same structure as Table 6, but is expanded with statistics on the largest network within each partition (Edges, $\gamma-$ and $\alpha-$ index).

Table 7: results of Thin Road Network algorithm

	total number of features		number of networks and statics on features in individual networks				Largest network		
Partition	Input	Result	COUNT	MIN	MAX	MEAN	Edges	γ – index	α - index
1	445	317	19	1	168	17	168	0.348	0.019
17	108	54	19	1	32	3	32	0.333	-0.016
49	29	14	5	1	8	3	8	0.444	0.091
93	820	766	16	1	711	48	711	0.396	0.094
198	1560	1408	13	1	1306	108	1306	0.369	0.053
217	137	94	6	1	53	16	53	0.327	-0.019
260	397	280	20	1	213	14	213	0.376	0.061
298	300	173	31	1	71	6	71	0.408	0.104
345	582	459	52	1	311	9	311	0.366	0.048
403	802	615	106	1	381	6	381	0.360	0.038
442	667	528	15	1	190	35	190	0.401	0.098

Data from this table can be compared with the data in Table 6 and shows significant decreases in number of networks for each partition (COUNT-column) and of number of features for the largest network (MAX-column)

A surprising aspect of this data is the occurrence of negative values for the $\alpha-$ index of two partitions. Analysis of the network and its features indicate errors in the topological network. The largest network of partition 17, for example, appears to consist of two networks. Caused by the computation of centrelines for water area features. The created centreline is not connected to existing water line features.

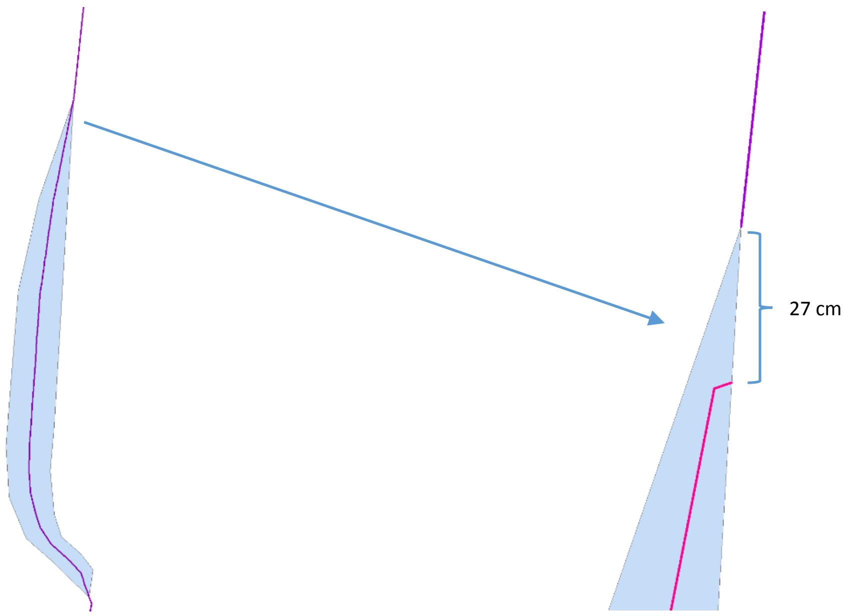


Figure 57: not-connected water centre line area causing gaps in largest network

The valid results will be used in a comparison of the results of the alternative thinning algorithms.

4.6.2 Stroke-based approach

The second thinning experiment, the stroke-based approach, is based on continuity of lines and is explained in full detail in paragraph 2.4.3.

4.6.2.1 Algorithm

The algorithm used in this experiment is based on the implementation in the research of Post (2014). In his master thesis, he researched stroke- and mesh-based thinning approaches in the context of road network generalisation. These algorithms have not been tested before in the context of thinning water networks. The models were adapted to fit the approach in this thesis. Figure 58 gives an overview of the algorithm. Appendix H can be consulted for full implementation.

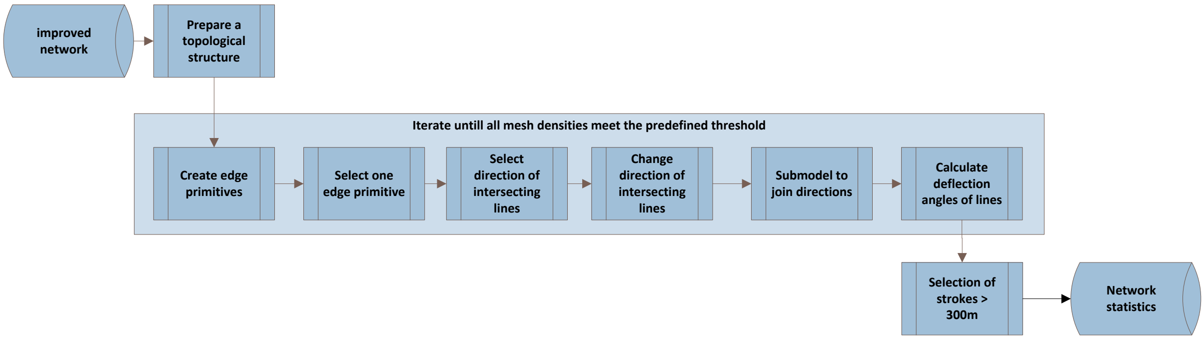


Figure 58: stroke-based algorithm

First, the algorithm creates a topological line structure consisting of nodes and edges to ensure the processing handles all features only once. Next, it runs iteratively over all edges and connects each edge to the neighbouring edge with the smallest deflection angle. The deflection angle between both edges must be smaller than 20 degrees for invoking a merge-action. The iteration stops when there are no more candidates which can be merged.

The stroke-based algorithm is in essence not a thinning algorithm. Rather it should be regarded as a selection algorithm to identify a geometry-based hierarchy for a network. In a subsequent processing action the real thinning must be implemented and results are dependent on the sophistication of such an algorithm (i.e. it should also take density and connectivity into account). For this research a straightforward selection on feature length is implemented.

4.6.2.2 Results of Stroke-based algorithm

The results of the stroke-based experiments for each partition are presented in Table 8. The table also includes statistics (number of edges, γ- and the α-index) for the largest network in each partition.

Table 8: results of Stroke-based algorithm

	total number of features		number of networks		number of features in individual networks			largest network		
partition	**input**	**results**	**input**	**results**	**min**	**max**	**mean**	**edges**	**γ – index**	**α - index**
1	445	168	118	37	1	24	4.541	6	0.50000	0.14286
17	108	52	68	21	1	26	2.476	19	0.35185	0.00000
49	29	8	15	5	1	3	1.600	1	0.00000	0.00000
93	820	496	58	58	1	101	8.552	81	0.42857	0.13600
198	1560	516	48	104	1	53	4.962	29	0.34524	0.00000
217	137	41	23	15	1	11	2.733	8	0.38095	0.00000
260	397	139	112	35	1	34	3.971	19	0.39583	0.06452
298	300	102	148	36	1	32	2.833	24	0.36364	0.02326
345	582	268	140	83	1	41	3.229	21	0.35000	0.00000
403	802	408	241	152	1	55	2.684	47	0.34815	0.01124
442	667	269	99	50	1	46	5.380	29	0.46032	0.17073

The MAX number (most features) does not equal the number of Edges for the largest network because the features are merged together to create a topological structure of edges and nodes. One edge can consist of multiple input features.

Zero values for respectively the γ – and the α –indices represent two phenomena. For the γ –index, a 0 value indicates a minimal connected network, consisting of a single edge. Networks with a γ –index value of 0 will always have an α –index value of 0, which implies there are no circuits. A zero value for the α -index implicates a network without any circuits. This is the implication of a network with multiple edges, but without alternative routes in the network (c.f. paragraph 2.6.1.2.2). This is illustrated by the results of the Stroke algorithm in Figure 59. Maps for all test areas are includes as “Strokes” in Figure 102 to Figure 112 in Appendix K.

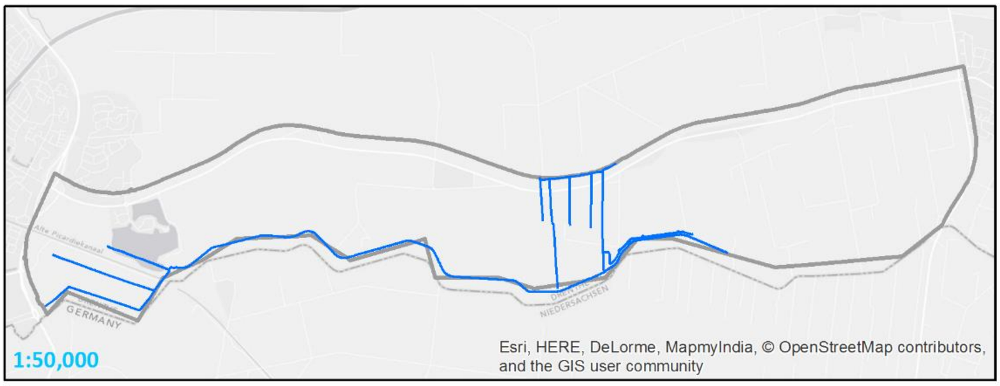


Figure 59: largest network without circuits

4.6.3 Mesh-based approach

The mesh-based approach is based on density of lines. The concept is explained in full detail in paragraph 2.4.4. The problem of overcrowded areas, which are identified on basis of a predetermined threshold value, is solved by removal of the longest shared edge with adjacent features.

4.6.3.1 Algorithm

The mesh-based approach starts with the results of the improve connectivity model (paragraph 4.5 and Appendix F). Figure 60 gives a conceptual overview.

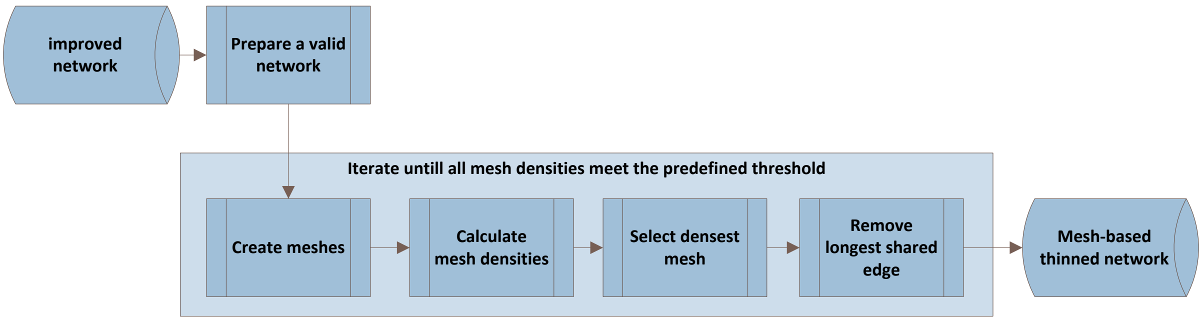


Figure 60: mesh-based algorithm

To prepare all data for processing, the first action in the algorithm is the construction of a topological structure with nodes and edge. Subsequently the edges are input to the mesh algorithm and meshes (polygons) are created. For each mesh the density is calculated and the densest mesh is selected. Based on the longest shared side for this mesh, an edge is removed from the topological structure. This mesh creation and edge removal-process is repeated until the densest mesh meets a predefined threshold density-value.

4.6.3.2 *Results of Mesh-based algorithm*

Table 9 presents the results obtained from the application of the mesh-based algorithm on each partition and statistics for the largest network in each partition. Negative values for the α –Index (partition 298 and 403) indicate broken, incomplete networks. These results are excluded from further analyses.

Table 9: results of Mesh-based algorithm

	total number of features		number of networks		number of features in individual networks			largest network		
partition	input	result	input	results	min	max	mean	edges	γ – index	α - index
1	445	445	118	338	1	49	4.364	16	0.444	0.130
17	108	100	68	63	1	84	3.746	32	0.344	0.000
49	29	26	15	14	2	27	5.143	7	0.583	0.285
93	820	789	58	162	1	227	10.920	100	0.388	0.076
198	1560	1482	48	228	1	823	13.658	348	0.406	0.107
217	137	133	23	52	1	52	5.788	24	0.381	0.049
260	397	389	112	168	1	71	4.988	31	0.356	0.017
298	300	288	148	190	1	113	3.363	51	0.327	-0.019
345	582	553	140	210	1	150	5.371	73	0.386	0.072
403	802	783	241	339	1	213	5.038	95	0.317	-0.030
442	667	661	99	179	1	296	7.905	132	0.411	0.113

The most surprising aspect of this data is the significant *increase* in the number of networks (with exception of partition 17), while hardly any thinning occurs (compare the number of resulting features with the number of input features). The reason for this is the thinning breaks existing networks in multiple parts, resulting in loss of continuing patterns.

4.7 Findings

As explained in the introduction, generalisation involves the abstracting of space, enabling the map reader to discern patterns and structures in topographical data (Kraak 2010). One of the tools to achieve generalisation is the *reduction* of detail in a map. However, this must be done in a meaningful way, to assure *resemblance* with structures in the input data and retain *connectivity* in the results. Along these three metrics the results of the thinning algorithms are evaluated: reduction of features, resemblance with input data and connectivity of features within the network.

The paragraphs below discuss the results for the three thinning algorithm in coherence. Based on the relative outcomes, the thinning algorithms will be ranked on average results for all partitions, but also for each individual partition, using the ranks 1, 2 and 3 (where 1 is the least important, while 3 is the most important). The scores will be evaluated in a simple multi-criteria analysis (MCA).

4.7.1 Reduction of features

The first criterion to compare and evaluate the results of the thinning algorithm alternatives for the different partitions is the reduction in features for the test areas. Table 10 and Figure 61 show the number of features for the input data (the prepared data of paragraph 4.5), and the results for the three algorithms. The reduction of features is compared to the input data for each thinning algorithm and is expressed as a relative decline. The largest negative percentage presents the largest decrease in feature count.

Table 10: results of thinning algorithm evaluated by feature amount reduction

partition	input	meshes	%	strokes	%	TRN	%	RANK
1	445	445	0%	168	-62%	317	-29%	STM
17	108	100	-7%	52	-52%	54	-50%	STM
49	29	26	-10%	8	-72%	14	-52%	STM
93	820	789	-4%	496	-40%	766	-7%	STM
198	1560	1482	-5%	516	-67%	1408	-10%	STM
217	137	133	-3%	41	-70%	94	-31%	STM
260	397	389	-2%	139	-65%	280	-29%	STM
298	300	288	-4%	102	-66%	173	-42%	STM
345	582	553	-5%	268	-54%	459	-21%	STM
403	802	783	-2%	408	-49%	615	-23%	STM
442	667	661	-1%	269	-60%	528	-21%	STM
		average change	**-4%**		**-60%**		**-29%**	**STM**

After comparison of the results in Table 10, the three algorithms can be ranked. The overall performance of the algorithm can be ranked as Stroke-based (3 points), TRN-based (2 points) and mesh-based (1 point). The same is true for ranking the individual partitions. However, it must be noted that the TRN-based algorithm shows more variance in the individual test areas, as shown in Figure 61. The application of other, more-hydrography based thinning algorithms could provide better results, but this is not part of this study.

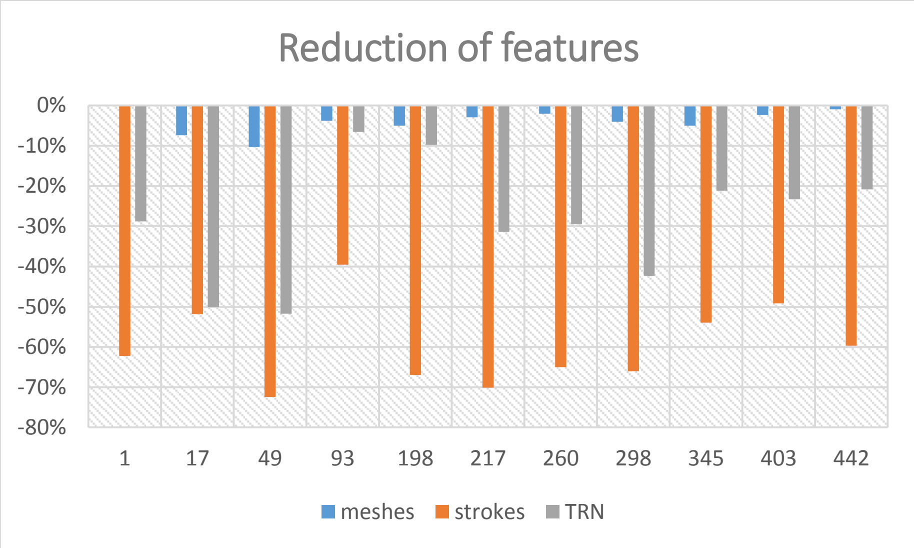


Figure 61: results of thinning algorithm evaluated by feature amount reduction

4.7.2 Resemblance

The second criterion for comparison of the outcomes of the alternative algorithms is resemblance with the input data. Resemblance can be defined as "X showing the same character as Y". It is not easy to find a metric which can be used for this, but the feature morphology can be used as an evaluation metric. The concept or the Feature Morphology Index (FMI) was introduced and applied in 4.2.2. FMI is a value expressing the relation between a number of features and the number of corresponding vertices. This can be used to evaluate average feature morphology within an area and can also be used for comparison between areas or to evaluate the effects of different approaches.

Table 11: comparison of change in average feature morphology per algorithm for each partition

partition	input	strokes	%	meshes	%	TRN	%	RANK
1	0.235	0.076	-68%	0.228	-3%	0.174	-26%	MTS
17	0.389	0.200	-49%	0.374	-4%	0.251	-35%	MTS
49	0.475	0.214	-55%	0.442	-7%	0.353	-26%	MTS
93	0.648	0.340	-47%	0.644	-1%	0.626	-3%	MTS
198	0.761	0.411	-46%	0.771	1%	0.745	-2%	MTS
217	0.643	S0.278	-57%	0.617	-4%	0.590	-8%	MTS
260	0.629	0.270	-57%	0.630	0%	0.558	-11%	MTS
298	0.517	0.228	-56%	0.542	5%	0.392	-24%	MTS
345	0.689	0.383	-44%	0.689	0%	0.637	-8%	MTS
403	0.675	0.444	-34%	0.672	0%	0.639	-5%	MTS
442	0.616	0.300	-51%	0.611	-1%	0.588	-5%	MTS
Average deviation in FMI			**-51%**		**-1%**		**-14%**	**MTS**

The FMIs of the mesh-based approach deviate the least from the input's FMIs. However, these figures are misleading since there is minimal reduction in amount of features (see the previous section). Input and output features are almost equal.

The stroke-based approach results in the most deviation and lowest FMI – values. This indicates the remaining features in these results are mainly winding. This can be explained by effect of the radical elimination of all features shorter than the threshold. All remaining features are longer than the threshold value and longer features are likely to be more winding than shorter features.

The results in the FMIs of the TRN-based approach reflect a middle position. Taking the feature reduction (on average 29%) into account these results reflect a good representation (an average deviation of 14%) of the morphology of the input data. In contrast to the results of the stroke-thinning algorithm, the TRN-based algorithm preserves connecting features which do not meet the length threshold value and seems to take connectivity (see the results in the next paragraph) into account.

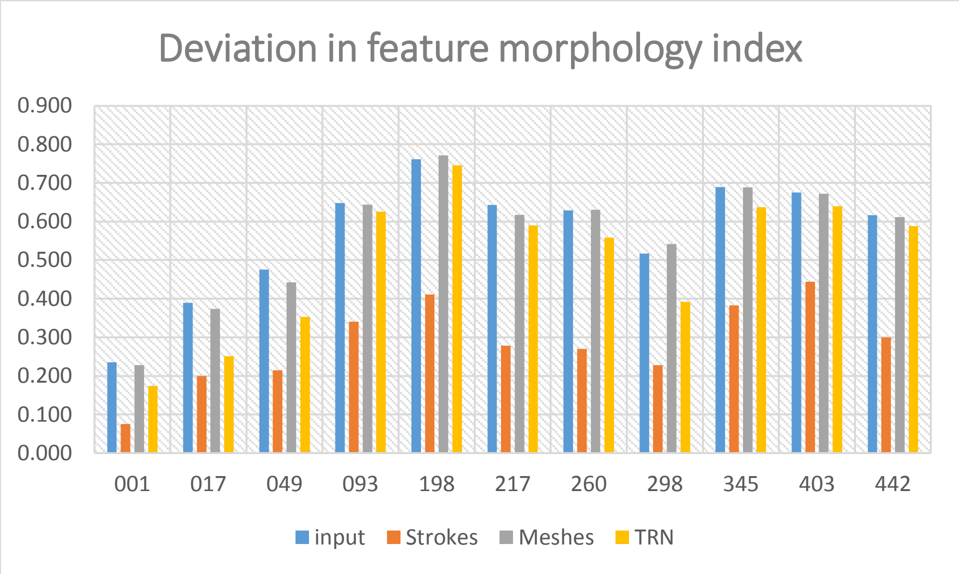


Figure 62: morphology index per algorithm per partition

In summary, the approaches can be ranks as: mesh-based (3 points), TRN-based (2 points) and stroke-based (1 point).

4.7.3 Connectivity

The third criterion to compare the results is the connectivity level of the features. At first glance using network statistics seems the logical choice. Unfortunately, these metrics can only be applied to one network at a time. For the comparison of the connectivity of the test partition it is not useful. For this reason a different metric is chosen for comparison, the reduction in number of networks.

Table 12: comparison of reduction in number of networks for each algorithm

	input	meshes	%	strokes	%	TRN	%	RANK
1	118	338	186%	37	-69%	19	-84%	TSM
17	68	63	-7%	21	-69%	19	-72%	TSM
49	15	14	-7%	5	-67%	5	-67%	TSM
93	58	162	179%	58	0%	16	-72%	TSM
198	48	228	375%	104	117%	13	-73%	TSM
217	23	52	126%	15	-35%	6	-74%	TSM
260	112	168	50%	35	-69%	20	-82%	TSM

298	148	190	28%	36	-76%	31	-79%	TSM
345	140	210	50%	83	-41%	52	-63%	TSM
403	241	339	41%	152	-37%	106	-56%	TSM
442	99	179	81%	50	-49%	15	-85%	TSM
average change			**+100%**		**-36%**		**-73%**	**TSM**

Comparing the results of stroke-based, mesh-based and TRN-based algorithms (Table 12), it can be seen that the most radical reduction in number of networks occurs when applying a TRN-based algorithm, while application of a mesh-based approach results in an *increase* of networks for most of the test partitions. The stroke-based algorithm results in the second best reduction, but is not always stable: for partition 93 the number of networks stays the same and for partition 198 the number of networks increase.

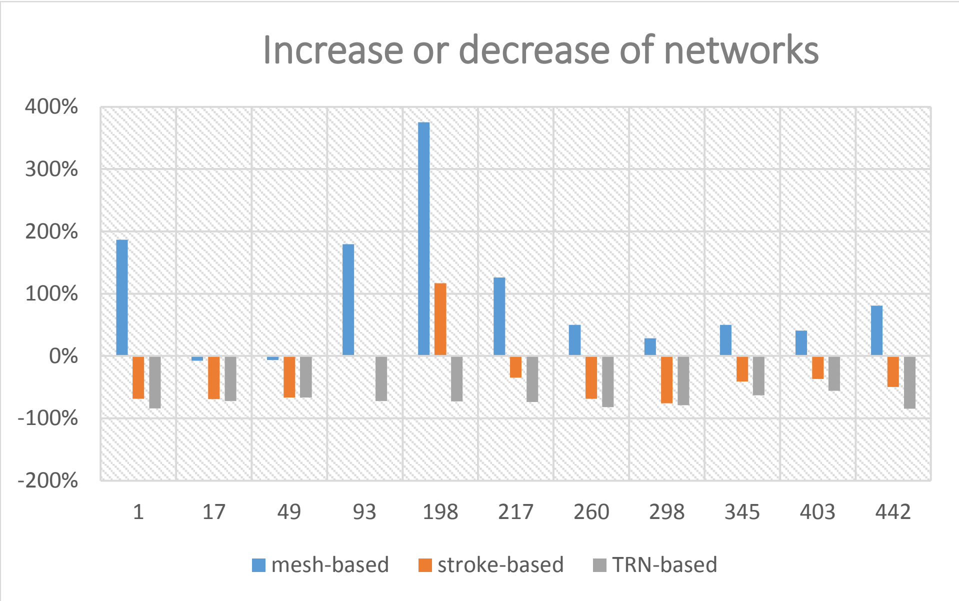


Figure 63: feature reduction per algorithm for each partition

Considering the connectivity-metric, results can be ranked as: TRN-based (3 points), stroke-based (2 points) and mesh-based (1 point).

4.7.4 Largest network analyses

The proposed evaluation metrics for networks statistics (γ- and α- indices) have been calculated for the largest network in each partition and are included in together with the visual results in Appendix K. However, these results are not useful for comparison because the visual results in Appendix K show that the largest network is not always located in the same geographical area (most evident in partition 1, 198 and 442); and because the extent of the largest network for each other partitions varies significantly (compare the number of edges within the largest networks in Table 7 to Table 9 and the results in Appendix K).

4.8 Overall conclusion and summary

The scores on the individual metrics (Reduction, Resemblance and Connectivity) for the three thinning algorithms together indicate the most suitable approach. The total score can be computed by accumulation of the individual scores. All metrics have equal weights in the MCA, because the three criteria are of equal importance.

	stroke-based	**mesh-based**	**TRN-based**
Reduction	3	1	2
Resemblance	1	3	2
Connectivity	2	1	3
Total	**6**	**5**	**7**

The results of the MCA indicate a clear hierarchy for the performance of algorithms: the Thin Road Network scores best, followed the by stroke-based algorithm. The results of the mesh-based algorithm should be considered not useful, because it scores are lowest for both reduction (an average decrease of -4%) and connectivity (an average increase of +100%).

The experiments in this chapter each addressed an aspect of the research questions, but it is important to notice all are part of a prototyping framework. The prototyping framework was divided in three stages, preparation, thinning and evaluation. Each subsequent stage was dependent upon the results of its predecessor, but it appeared to be possible to measure intermediate results. The individual results as well as all results in coherence provided insight of the impact of preparatory work on thinning experiments. Also the influence of landscape typology on thinning algorithm was measured, but this did not provide significant deviation in comparison with the average results.

However, the variance in the feature-reduction results of the TRN-based method could be basis for investigation of other thinning algorithms (more dedicated to hydrography).

5 Conclusion

This research was conceived during the design of automatic generalisation of a topographic map series. Results of this generalisation were unsatisfactory in appreciation of landscape characteristics and in consistency and coherence of generalisation results. To study and tackle these issues, the present study was designed to deliver a methodology for pruning of artificial manmade networks with regard to landscape typology and to research methods for evaluation of the quality of generalisation results.

5.1 Overview of results

The first question, "What is a suitable methodology for pruning artificial manmade network which takes landscape typology in account?", addresses the concept of methodology on tailoring thinning algorithms for networks, based on landscape typology. The question is answered by the introduction of a prototyping framework (4.1), which offers methods to identify landscape type for network features (4.2) and to improve connectivity within a network (4.5).

The identification of landscape type for an area was researched using a hybrid approach (4.2). An existing landscape classification dataset has been evaluated on its merits to show variance in feature morphology and humidity. Results showed enough variety in the landscape for use in the selection of test areas (4.3). This experiment also revealed the usability of morphology (4.2.2) and humidity (4.2.1) as metrics to characterise an area.

It was also shown that the connectivity of a network can be improved by adding existing geometries and computed perceptual links to the dataset (4.5). This results in reduction of number of networks. In conclusion, the results of this algorithm show significant improvement in network connectivity. However, it appears that network improvement still allows the existence of isolated features (4.5.3).

The usage and behaviour of this enhanced network in alternative thinning algorithms was subject of the next set of experiments (4.6). The application of three thinning algorithms have been researched: a stroke-based approach, a mesh-based approach and a TRN-based approach.

The stroke-based algorithm (4.6.2) is in essence not a thinning algorithm. Rather it should be regarded as a selection algorithm to identify a geometry-based hierarchy for a network. In a subsequent processing action the real thinning must be implemented and results are dependent on the sophistication of such an algorithm (i.e. it should also take density and connectivity into account). (4.6.2.1)

The most significant aspect in the results of the mesh-based algorithm (4.6.3) is the *increase* in the number of networks, while hardly any thinning occurs. The reason for this is the thinning breaks existing networks in multiple parts, resulting in loss of continuing patterns. (4.6.3.2)

The results of the TRN-based approach (4.6.1) show to be most promising both in general and in the test scores for the individual areas (4.6.1.2).

The second research question "How can thinning results be evaluated objectively?" centred on the concept of Evaluation. In the Literature study the concepts and relevance of existing evaluation algorithms have been discussed (2.6). It appeared existing evaluation methods have limited application possibilities for this study, due to requirements on benchmark data (2.6.1.1) or network structures (2.6.1.2 and 4.7.4). Therefore new three evaluation metrics on Reduction, Resemblance and Connectivity were designed for use in comparison of alternative thinning approaches (4.7).

In a comparison of the results of the three approaches an evaluation was executed based on the coherence of these metrics. Each test area was scored on its performance and the ranking was decided by using a Multi Criteria Analysis (4.8). The results of the MCA indicate a clear hierarchy for the performance of algorithms: the Thin Road Network scores best, followed the by stroke-based algorithm. The results of the mesh-based algorithm should be considered not useful, because it scores are lowest for both reduction and connectivity.

5.2 Limitations and recommendations for further research

Although the study has successfully demonstrated the identification of landscape typology and its use for evaluation purposes, it has provided limited results in tailoring the choice for the used thinning algorithms with regard to landscape type. It is possible to distinguish landscape type by humidity and feature morphology and to use this to tailor the pruning algorithm, but the researched thinning algorithms did not provide significant differences in results on the individual landscape types (4.7).

It would be interesting to assess the effects of different thinning algorithms on the identified landscape types (i.e. (Savino 2014; Brewer et al. 2009; Buttenfield 2010). The proposed framework provides room for inserting and comparison of different thinning solutions.

A future study investigating data enrichment from external data sources to acquire a more sophisticated hierarchy for thinning purposes and its effect to generalisation results would be very interesting,

References

Altena, V. van et al., 2013. Complete generalisation of 1:50k roads centrelines from 1:10k road polygons. In 16 th ICA Workshop on Generalisation and Map Production. Dresden, Germany.

Anderson-Tarver, C. et al., 2011. Automated delineation of stream centerlines for the USGS National Hydrography Dataset. In *Advances in Cartography and GIScience. Volume 1*. Springer, pp. 409–423. Available at: http://link.springer.com/chapter/10.1007/978-3-642-19143-5_23 [Accessed February 5, 2014].

ArcGIS Help 10.2, 2014. Thin Road Network (Cartography). Available at: http://resources.arcgis.com/en/help/main/10.2/index.html#/Thin_Road_Network/007000000014000000/ [Accessed November 18, 2014].

Baella, B. & Pla, M., 2014. Deriving Products Through Generalisation at Institut Cartogràfic de Catalunya. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 333–339.

Baella, B. & Pla, M., 2005. Reorganizing the Topographic Databases of the Institut Cartogràfic de Catalunya applying generalization. In 8th ICA WORKSHOP on Generalization and Multiple Representation, A Coruña, 7 th -8 th July, 2005. Spain. Available at: http://generalisation.icaci.org/images/files/workshop/workshop2005/Baella_Pla.pdf.

Bakker, N., Bruns, B. & Storm, M., 2005. Gegevensmodel TOP10NL versie 2.3.

Bakker, N. & Kolk, B., 2003. TOP10NL, a new object-oriented topographical database in GML. In La Coruña: Sociedad Española de Cartografía Fotogrametría y Teledetección.

Barrault, M. et al., 2001. Integrating multi-agent, object-oriented and algorithmic techniques for improved automated map generalization. In *Proceedings 20th International Cartographic Conference*. pp. 2110–2116. Available at: http://icaci.org/files/documents/ICC_proceedings/ICC2001/icc2001/file/f13041.pdf [Accessed February 5, 2014].

Benz, S. & Weibel, R., 2013. Road Network Selection Using an Extended Stroke-Mesh Combination Algorithm. In ICA Workshop on Generalisation and Map Production. Dresden, Germany, p. 11.

Bildirici, I.O., 2004. Building and Road Generalization with the CHANGE Generalization Software Using Turkish Topographic Base Map Data. *Cartography and Geographic Information Science*, 31(1), pp.43–54.

Brewer, C.A., Buttenfield, B.P. & Usery, E.L., 2009. Evaluating generalizations of hydrography in differing terrains for The National Map of the United States. In *Proceedings, 24 th International Cartographic Congress*. Available at: http://www.personal.psu.edu/users/c/a/cab38/ScaleMaster/Brewer_etal_ICC2009_paper_Nov2009.pdf [Accessed February 5, 2014].

Bruns, B. et al., 2010. *Automatische generalisatie TOP10NL naar TOP50Actueel - eindrapportage High5 project*, Zwolle: Kadaster.

Burghardt, D., Duchêne, C. & Mackaness, W.A. eds., 2014. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of Map Generalisation*, Cham: Springer International Publishing. Available at: http://nbn-resolving.de/urn:nbn:de:1111-20140228111 [Accessed March 28, 2014].

Buttenfield, B.P. et al., 2013. Automatic Enrichment of Stream Networks with Primary Paths for Use in the United States National Atlas. Available at: http://icaci.org/files/documents/ICC_proceedings/ICC2013/_extendedAbstract/238_proceeding.pdf [Accessed February 5, 2014].

Buttenfield, B.P., 2010. Generalization of Hydrographic Features and Automated Metric Assessment Through Bootstrapping.

Buttenfield, B.P., Stanislawski, L.V. & Brewer, C.A., 2010. Multiscale representations of water: Tailoring generalization sequences to specific physiographic regimes. *Proceedings of GIScience 2010*, pp.14–17.

Chaudhry, O. & Mackaness, W., 2005. Rural and Urban Road Network Generalisation Deriving 1:250,000 from OS MasterMap. In S. E. de C. F. y T. International Cartographic Conference, ed. *Abstracts of papers: International Cartographic Conference, 9-16 July, 2005, A Coruña, Spain.* [La Coruña]: Sociedad Española de Cartografía Fotogrametría y Teledetección.

Chen, J. et al., 2009. Selective omission of road features based on mesh density for automatic map generalization. *International Journal of Geographical Information Science*, 23(8), pp.1013–1032.

Cultural Heritage Agency of the Netherlands, 2014. Cultural Heritage Agency of the Netherlands | Rijksdienst voor het Cultureel Erfgoed. *Cultural Heritage Agency of the Netherlands*. Available at: http://www.culturalheritageagency.nl/en [Accessed October 21, 2014].

Douglas, D.H. & Peucker, T.K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2), pp.112–122.

Duchêne, C. et al., 2014. Generalisation in Practice Within National Mapping Agencies. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 329–392.

Duchêne, C., Ruas, A. & Cambier, C., 2012. The CartACom model: transforming cartographic features into communicating agents for cartographic generalisation. *International Journal of Geographical Information Science*, 26(9), pp.1533–1562.

Edwardes, A. & Mackaness, W., 2000. Intelligent road network simplification in urban areas. *Available on http://www. geos. ed. ac. uk/homes/wam/Edwardes-Mack2000. pdf*. Available at: http://www.geos.ed.ac.uk/homes/wam/EdwardesMack2000.pdf [Accessed February 5, 2014].

Farjon, J.M.J. et al., 2001. *Neder-landschap Internationaal: bouwstenenen voor een selectie van gebieden landschapsbehoud*, Wageningen: Alterra Research Instituut voor de Groene Ruimte.

Galanda, M., 2003. Automated polygon generalization in a multi agent system. *Unpublished Ph. D. Dissertation, Department of Geography, University of Zurich, Switzerland*. Available at: http://www.geo.uzh.ch/fileadmin/files/content/abteilungen/gis/research/phd_theses/thesis_MartinGalanda_2003.pdf [Accessed February 5, 2014].

Garnero, G. & Godone, D., 2012. Generalized map production: Italian experiences. In *Knowing to manage the territory, protect the environment, evaluate the cultural heritage*. FIG Working Week 2012. Rome, Italy, p. 9.

Gould, N., 2014. *Formalising cartographic generalisation knowledge in an ontology to support on-demand mapping*. Manchester: Manchester Metropolitan University.

Grünreich, D., 1988. Untersuchungen zum Aufbau von ATKIS in Niedersachsen. *Nachrichten aus dem Karten- und Vermessungswesen*, 3(38).

Han, H. & Qiao, J., 2010. A Self-Organizing Fuzzy Neural Network Based on a Growing-and-Pruning Algorithm. *IEEE Transactions on Fuzzy Systems*, 18(6), pp.1129–1143.

Hardy, P., Lee, D. & van Smaalen, J., 2008. Practical research in generalization of european national framework data from 1: 10K to 1: 50K, exercising and extending an industry-standard GIS. In *ICA Workshop on Generalization and Multiple Representation, Montpellier. http://aci. ign. fr/montpellier2008/papers/19_Hardy_et_al. pdf*. Available at: http://pghardy.net/paul/papers/2008_icagmr_montpellier_hardy_lee_smaalen.pdf [Accessed February 5, 2014].

Harrie, L. & Weibel, R., 2007. Modelling the Overall Process of Generalisation. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski, eds. *Generalisation of geographic information cartographic modelling and applications*. Amsterdam; Boston: Published on behalf of the International Cartographic Association by Elsevier, pp. 67–87.

Heinzle, F., Anders, K.H. & Sester, M., 2005. Graph based approaches for recognition of patterns and implicit information in road networks. In *Proceedings of 22nd international cartographic conference, A Coruna, Spain*. Available at: http://icaci.org/files/documents/ICC_proceedings/ICC2005/htm/pdf/oral/TEMA9/Session%204/FRAUKE%20HEINZLE.pdf [Accessed February 5, 2014].

Horace: (22-8 BC), 8AD. Ars Poetica, 343. Available at: http://www.thelatinlibrary.com/horace/arspoet.shtml [Accessed April 25, 2014].

ICA Commission on Generalisation and Multiple Representation, 2013. NMA Symposium 2013 - General information. *NMA Symposium - General information*. Available at: http://generalisation.icaci.org/index.php/nextevents/12-next-events-details/48-nma-symposium-2013-general [Accessed August 25, 2014].

International Cartographic Association - Commission II, 1973. *Multilingual Dictionary of Technical Terms in Cartography; Dictionnaire Multilingue De Termes Techniques Cartographiques: Diccionario Multilingue De Terminos Tecnicos Cartograpficos: Mehrsprachiges Worterbuch Kartographischer Fachbegriffe*, Wiesbaden, Germany: Franz Steiner Verlag.

Jäger, E., 1991. Investigations on automated feature displacement for small scale maps in raster format. *Proceedings, ACI*, pp.245–256.

Jiang, B. & Claramunt, C., 2004a. A structural approach to the model generalization of an urban street network*. *GeoInformatica*, 8(2), pp.157–171.

Jiang, B. & Claramunt, C., 2004b. Topological analysis of urban street networks. *Environment and Planning B: Planning and Design*, 31(1), pp.151–162.

Jiang, B., Zhao, S. & Yin, J., 2008. Self-organized natural roads for predicting traffic flow: a sensitivity study. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(07), p.P07008.

Kadaster, 2013. Basisregistratie Topografie: catalogus en specificaties.

Käuferle, D., 2014. Producing Digital Cartographic Models at Swisstopo. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 346–351.

Kraak, M.J., 2010. *Cartography: visualization of geospatial data* 3rd ed., Harlow ; New York: Prentice Hall.

Lamy, S. et al., 1999. The application of agents in automated map generalisation. In *19th international cartographic conference*. Available at: http://www.geos.ed.ac.uk/homes/wam/Agent-ICA1999.pdf [Accessed March 14, 2014].

Lecordix, F. et al., 2007. Development of a new generalization flowline for topographic maps. In *10th ICA workshop on Generalisation and Multiple Representation*. pp. 2–3.

Lecordix, F. & Maugeais, E., 2014. The New Base Map Project: A Semi-Automated Production Line for Topographic Maps at IGN-France. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 339–346.

Liu, X., Zhan, F.B. & Ai, T., 2010. Road selection based on Voronoi diagrams and “strokes” in map generalization. *International Journal of Applied Earth Observation and Geoinformation*, 12, pp.S194–S202.

Li, Z. & Zhou, Q., 2012. Integration of linear and areal hierarchies for continuous multi-scale representation of road networks. *International Journal of Geographical Information Science*, 26(5), pp.855–880.

Luan, X. & Yang, B., 2010. Generating strokes of road networks based on pattern recognition. In *13th Workshop of the ICA Commission on Generalisation and Multiple Representation, Zurich, Switzerland*. p. 13. Available at: http://www.researchgate.net/publication/228562636_Generating_Strokes_of_Road_Networks_Based_on_Pattern_Recognition/file/60b7d51a6b13b13bb5.pdf [Accessed March 14, 2014].

Mackaness, W.A. & Beard, K.M., 1993. Use of graph theory to support map generalization. *Cartography and Geographic Information Systems*, 20(4), pp.210–221.

Mackaness, W.A., Ruas, A. & Sarjakoski, L.T., 2007. *Generalisation of geographic information cartographic modelling and applications*, Amsterdam; Boston: Published on behalf of the International Cartographic Association by Elsevier.

Makhzoumi, J., 1999. *Ecological landscape design and planning: the Mediterranean context*, London ; New York: E & FN Spon.

Meijers, M., 2011. *Variable-scale geo-information*, Delft: NCG.

Monnot, J.-L., Hardy, P. & Lee, D., 2007a. An optimization approach to constraint-based generalization in a commodity GIS framework. *ICC2007, Moscow, August*. Available at: http://pghardy.net/paul/papers/2007_icc_moscow_monnot_hardy_lee.pdf [Accessed February 5, 2014].

Monnot, J.-L., Hardy, P. & Lee, D., 2007b. Topological Constraints, Actions and Reflexes, for Generalization by Optimization. In *ICA Workshop on Generalisation and Multiple Representation, Moscow*. Available at: http://pghardy.net/paul/papers/2007_ica_moscow_monnot_hardy_lee_workshop.pdf [Accessed February 5, 2014].

Muller, J.C., 1991. Generalisation of Spatial Databases. In D. J. Maguire & M. F. Goodchild, eds. *Geographical information systems : principles and applications*. London: Longman, pp. 457–475.

Van Oosterom, P. et al., 2008. Een variabele-schaal structuur voor IMGeo en Top10NL. *GEO-INFO*, (11), pp.406–410.

Van Oosterom, P. & Meijers, M., 2011. Towards a true vario-scale structure supporting smooth-zoom. In *Proceedings of 14th ICA/ISPRS workshop on generalisation and multiple representation, Paris*. pp. 1–19. Available at: http://www.gdmc.nl/publications/2011/True_vario-scale_structure.pdf [Accessed February 5, 2014].

Oude Elberink, S.J., 2010. *Acquisition of 3D topography: automated 3D road and building reconstruction using airborne laser scanner data and topographic maps*. S.l.: s.n.].

Oxford dictionary, 2014. topography: definition of topography in. *Oxford dictionary (British & World English)*. Available at: http://www.oxforddictionaries.com/definition/english/topography [Accessed August 28, 2014].

Peng, W., 1997. *Automated generalization in GIS*. Enschede, the Netherlands: ITC.

Post, J.M., 2014. *Thinning of road network in the Dutch automatic generalisation*. Master Thesis. Manchester: The Manchester Metropolitan University.

Punt, E. & Watkins, D., 2010. User-Directed Generalization of Roads and Buildings for Multi-Scale Cartography. In *Proceedings of 13th ICA Workshop on Generalisation and Multiple Representation*. pp. 12–13.

Regnauld, N., 2014. Automatic Map Derivation at Ordnance Survey GB. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World*

Methodologies and Applications of Map Generalisation. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 351–355.

Regnauld, N. et al., 2014. Process Modelling, Web Services and Geoprocessing. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 197–225.

Revell, P., Regnauld, N. & Bulbrooke, G., 2011. OS VectormapTM District: Automated Generalisation, Text Placement and Conflation in support of making pubic data public. In *Proceedings of the 25th International Cartographic Conference Paris, 3-8 July 2011.* Proceedings of the 25th International Cartographic Conference Paris, 3-8 July 2011. Paris. Available at: http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/D3-Generalisation/CO-358.pdf.

Revell, P., Regnauld, N. & Thom, S., 2005. Generalising OS Mastermap® Topographic buildings and itn Road centerlines to 1: 50 000 scale using a spatial hierarchy of agents, triangulation and topology. In *Proceedings of the International Cartographic Conference*. Available at: http://icaci.org/files/documents/ICC_proceedings/ICC2005/htm/pdf/oral/TEMA9/Session%204/PATRICK%20REVELL.pdf [Accessed February 5, 2014].

Rodrigue, J.-P., 2013. *The geography of transport systems* Third edition., Abingdon, Oxon: Routledge.

Savino, S., 2014. Generalization of braided streams. In 17th ICA Workshop on Generalisation and Multiple Representation. Vienna, Austria.

Schmidt, M., 2014. International Cartographic Association - Mission. *International Cartographic Association*. Available at: http://icaci.org/mission/ [Accessed April 3, 2014].

Stanislawski, L.V., 2008a. Development of a knowledge-based network pruning strategy for automated generalisation of the United States National Hydrography Dataset. In *11th Workshop of the ICA commission on Generalisation and Multiple Representation, Montpellier, France*. pp. 20–21. Available at: http://pdf.aminer.org/000/295/196/algorithms_for_the_hydrographic_network_names_association_on_geographic_maps.pdf [Accessed February 5, 2014].

Stanislawski, L.V., 2009. Feature pruning by upstream drainage area to support automated generalization of the United States National Hydrography Dataset. *Computers, Environment and Urban Systems*, 33(5), pp.325–333.

Stanislawski, L.V., Buttenfield, B.P., et al., 2014. Generalisation Operators. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 157–196.

Stanislawski, L.V., 2008b. Intelligent Pruning of the United States National Hydrography Dataset. *Retrieved December*, 9, p.2009.

Stanislawski, L.V., Brewer, C.A. & Buttenfield, B.P., 2014. Generalisation Methods Used for the USGS National Map and National Atlas. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of*

Map Generalisation. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 351–355.

Stanislawski, L.V., Buttenfield, B.P. & Samaranayake, V.A., 2010. Generalization of Hydrographic Features and Automated Metric Assessment Through Bootstrapping. In 13th Workshop of the International Cartographic Association commission on Generalisation and Multiple Representation. Zurich, Switzerland,.

Stansilawski, L.V. et al., 2009. Stratified database pruning to support local density variations in automated generalization of the United States National Hydrography Dataset. In *Proceedings of 24 th International Cartographic Conference (ICC2009), Santiago Chile*. Available at: http://w.icaci.org/files/documents/ICC_proceedings/ICC2009/html/nonref/19_11.pdf [Accessed February 5, 2014].

Stoter et al., 2012. Automatische generalisatie voor nieuwe productielijn Kadaster. *GEO-INFO*, (2), pp.6–10.

Stoter, J., 2010a. Automatische generalisatie van topografische data. *GEO-INFO*, (2), pp.18–24.

Stoter, J., Zhang, X., et al., 2014. Evaluation in Generalisation. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 259–298.

Stoter, J. et al., 2011. Feasibility study on an automated generalisation production line for multiscale topographic products. Available at: http://www.gdmc.nl/publications/2011/Automated_generalisation_production_line.pdf [Accessed March 14, 2014].

Stoter, J., Post, M., et al., 2014. Fully automated generalization of a 1:50k map from 1:10k data. *Cartography and Geographic Information Science*, 41(1), pp.1–13.

Stoter, J. et al., 2012. Implementation of a National 3D standard: Case of the Netherlands. In J. Pouliot et al., eds. *Progress and New Trends in 3D Geoinformation Sciences*. Springer, p. 277–.

Stoter, J. et al., 2009. Specifying Map Requirements for Automated Generalization of Topographic Data. *The Cartographic Journal*, 46(3), pp.214–227.

Stoter, J., 2010b. *State-of-the-Art of Automated Generalisation in Commercial Software*, EuroSDR.

Stoter, J.E. et al., 2012. Fully automated generalisation of topographic data in current geo-information environments. *Urban and Regional Data Management–UDMS Annual 2012*, pp.111–121.

Stoter, J. & Smaalen, J.W.N. van, 2008. Automatische generalisatie van TOP10NL naar TOP50NL haalbaar? *GEO-INFO*, 2, pp.48–53.

Thomson, R.C. & Brooks, R., 2000. Efficient generalization and abstraction of network data using perceptual grouping. In *Proceedings of the 5th International Conference on GeoComputation*.

pp. 23–25. Available at: http://www.geocomputation.org/2000/GC029/Gc029.htm [Accessed February 5, 2014].

Thomson, R. & Richardson, D., 1999. *The "good continuation" principle of perceptual organization applied to the generalization of road networks*, ICA/ACI.

Travanca Lopes, J.A. & Catalão, C., 2002. A new method for line generalization based on artificial intelligence algorithms. In A. Ruas, ed. *Généralisation et représentation multiple*. Paris: Hermès science ; Lavoisier.

Urbanke, S. & Wiedemann, A., 2014. AdV-Project "Atkis: Generalisation"- Map Production of DTK50 and DTK100 at LGL in Baden-Württemberg. In D. Burghardt, C. Duchêne, & W. A. Mackaness, eds. *Abstracting Geographic Information in a Data Rich World Methodologies and Applications of Map Generalisation*. Lecture Notes in Geoinformation and Cartography. Cham: Springer International Publishing, pp. 369–373.

Wang, W.-C., 2011. Developing a line-of-sight based algorithm for urban street network generalization. *Applied Geomatics*, 4(2), pp.67–74.

Weiss, R. & Weibel, R., 2013. Road Network Selection for Small-Scale Maps Using an Improved Centrality Approach. In ICA Workshop on Generalisation and Map Production. Dresden, Germany, p. 11.

West-Nielsen, P. & Meyer, M., 2007. Automated Generalisation in a Map Production Environnement - the KMS Experience. In W. A. Mackaness, A. Ruas, & L. T. Sarjakoski, eds. *Generalisation of geographic information cartographic modelling and applications*. Amsterdam; Boston: Published on behalf of the International Cartographic Association by Elsevier, pp. 301–313.

Winkel, D. ter, 2012. Verkenningsvoorschriften versie 2012.1.

Wong, D.W.S., 2005. *Statistical analysis of geographic information with ArcView GIS and ArcGIS* Fully rev. & updated., Hoboken, N.J: John Wiley & Sons, Inc.

Yan, H., Li, Z. & Ai, T., 2006. System for automatic generalization of topographic maps. *Chinese Geographical Science*, 16(2), pp.165–170.

Zhou, Q. & Li, Z., 2012. A comparative study of various strategies to concatenate road segments into strokes for map generalization. *International Journal of Geographical Information Science*, 26(4), pp.691–715.

Appendices

Appendix A. Calculate humidity - algorithm

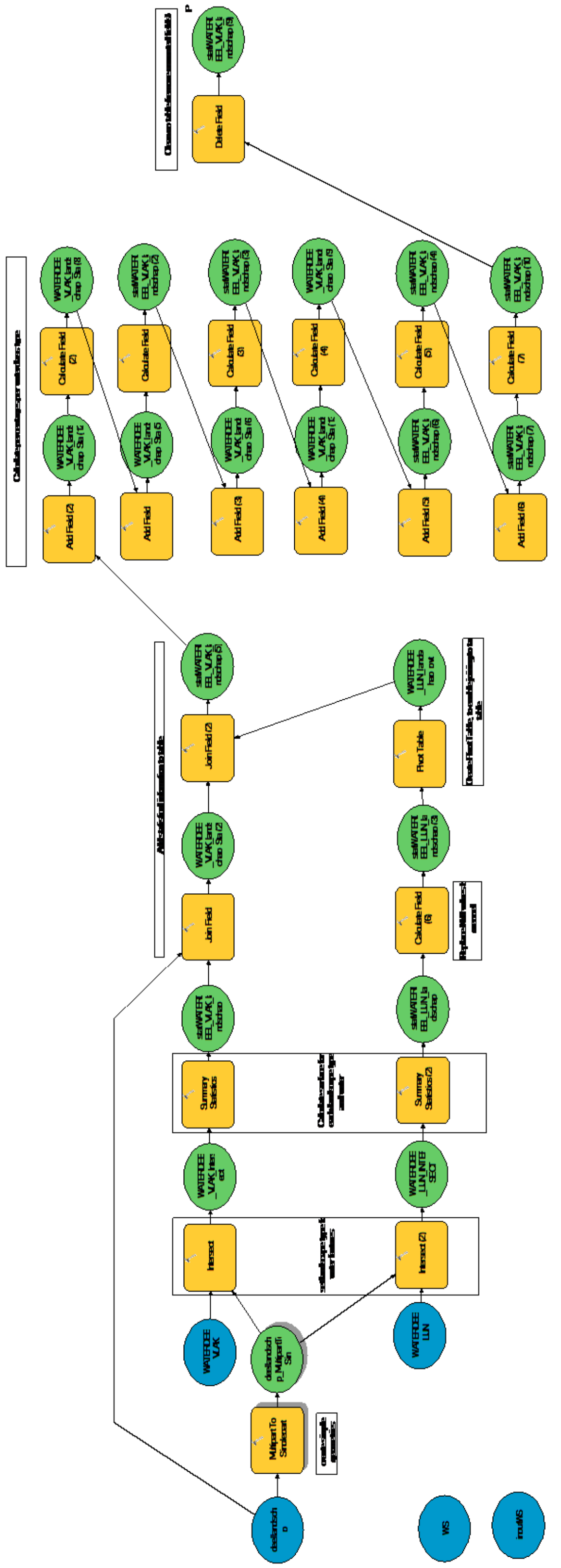


Figure 64: model to calculate humidity for each landscape type

Appendix B. Feature morphology algorithms

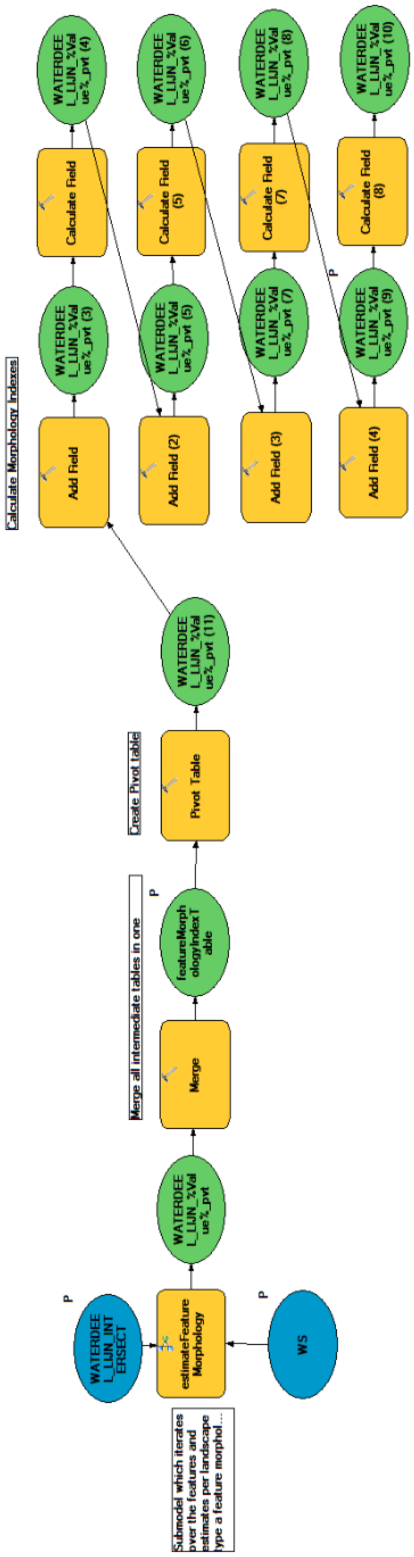


Figure 65: main morphology index model which finalizes the computation of statistics

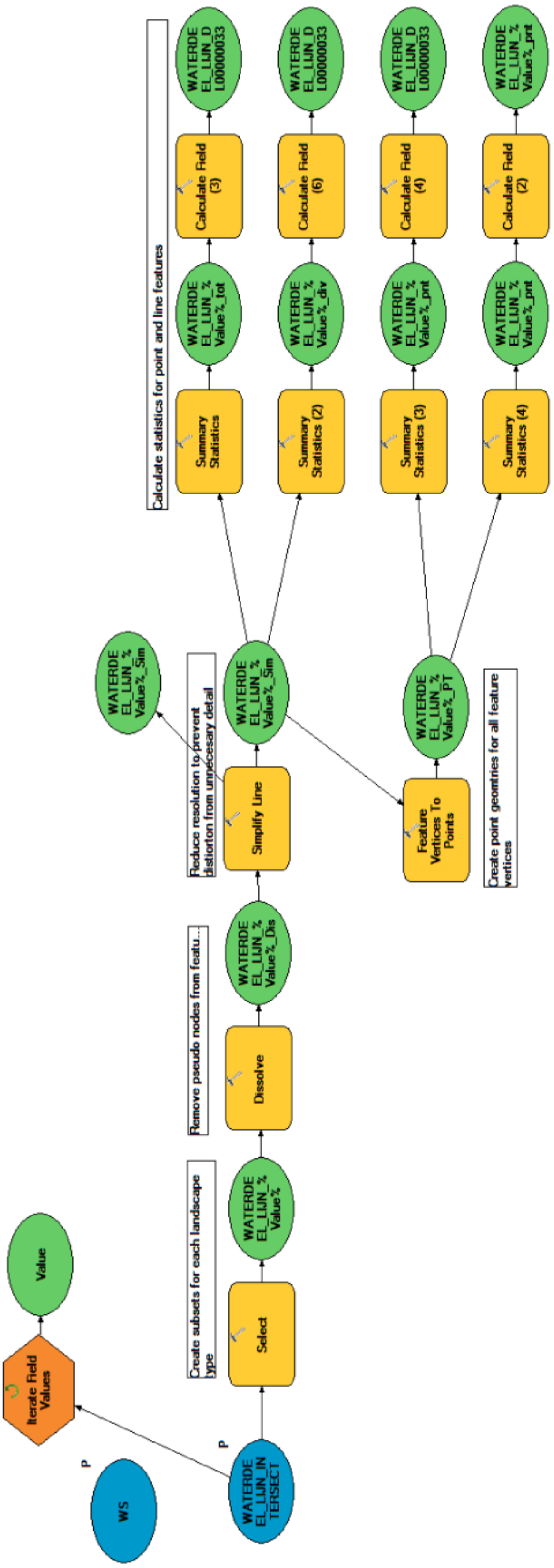


Figure 66: sub model for morphology index, computes statistics iterating over landscape types.

Appendix C. Selection of test partitions

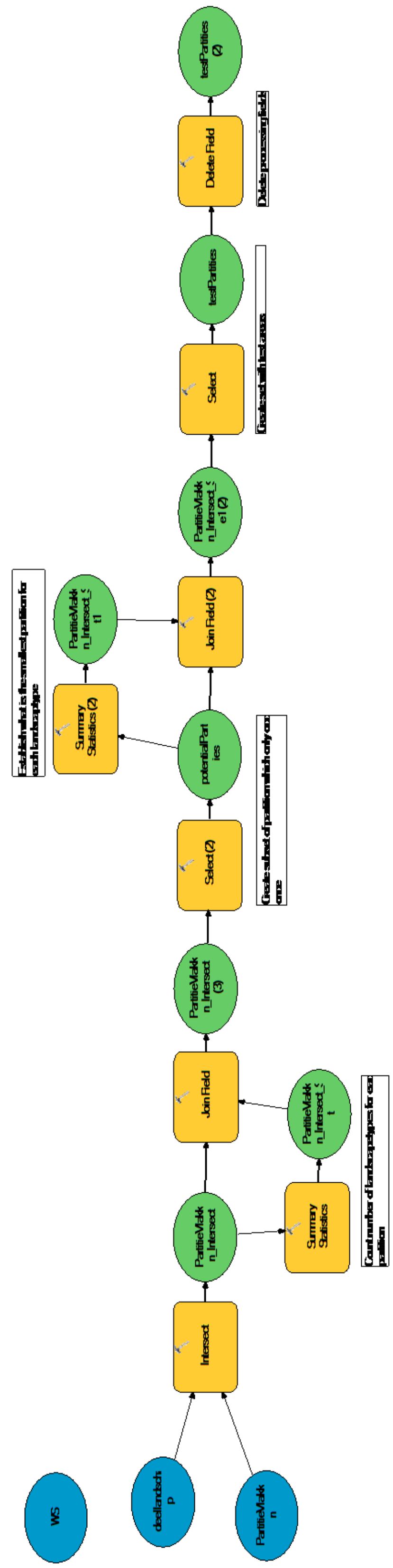


Figure 67: selection of test partitions

Appendix D. Count number of networks – algorithm

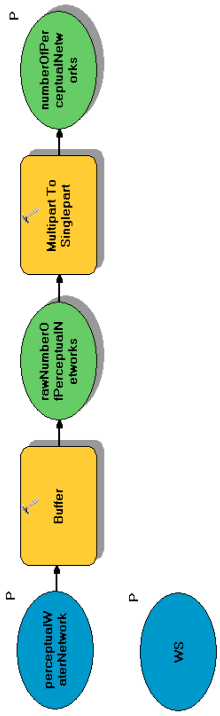


Figure 68: number of networks

Appendix E. Network statistics – algorithm

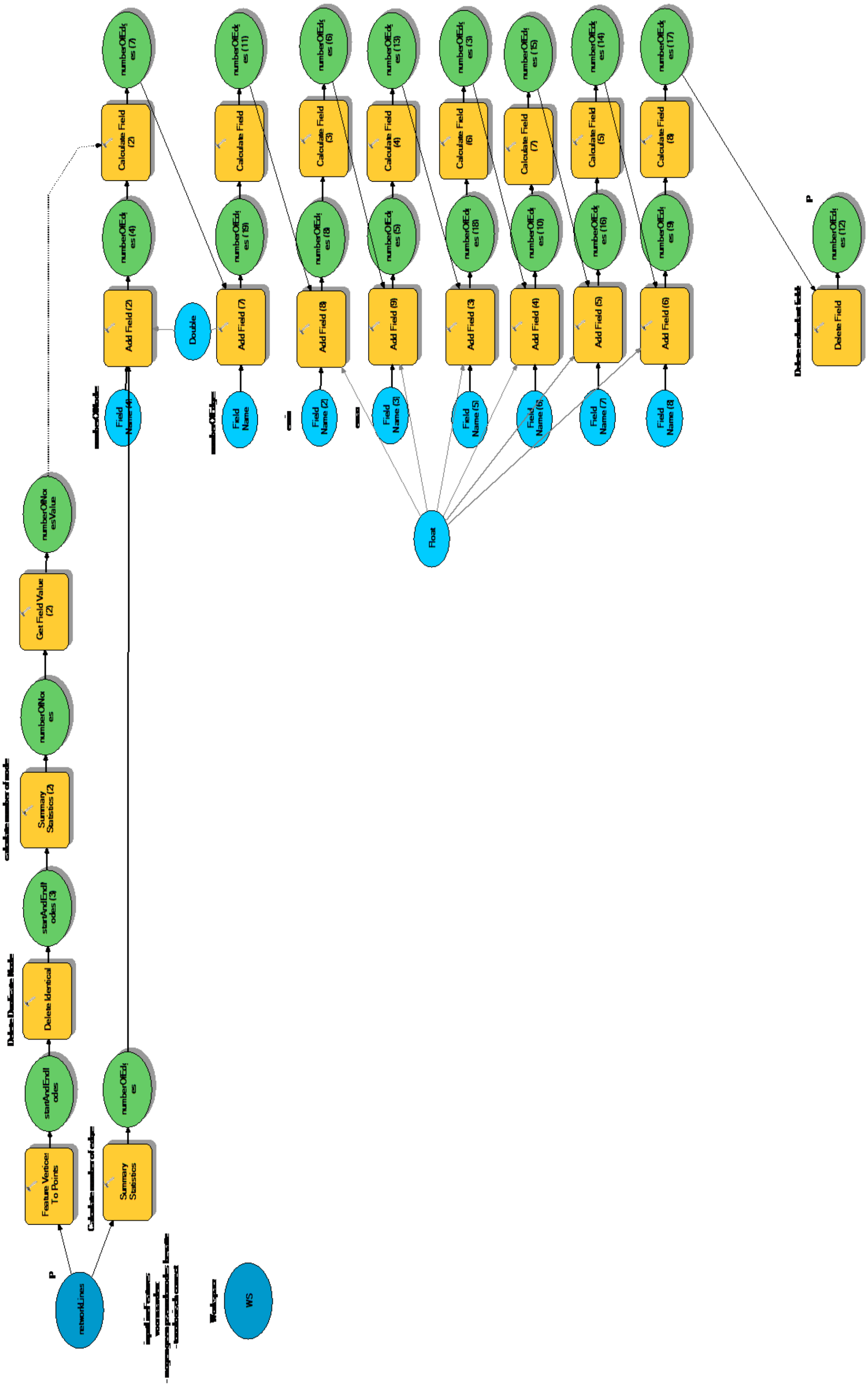


Figure 69: model to calculate connectivity indices

Appendix F. Improve network connectivity algorithm

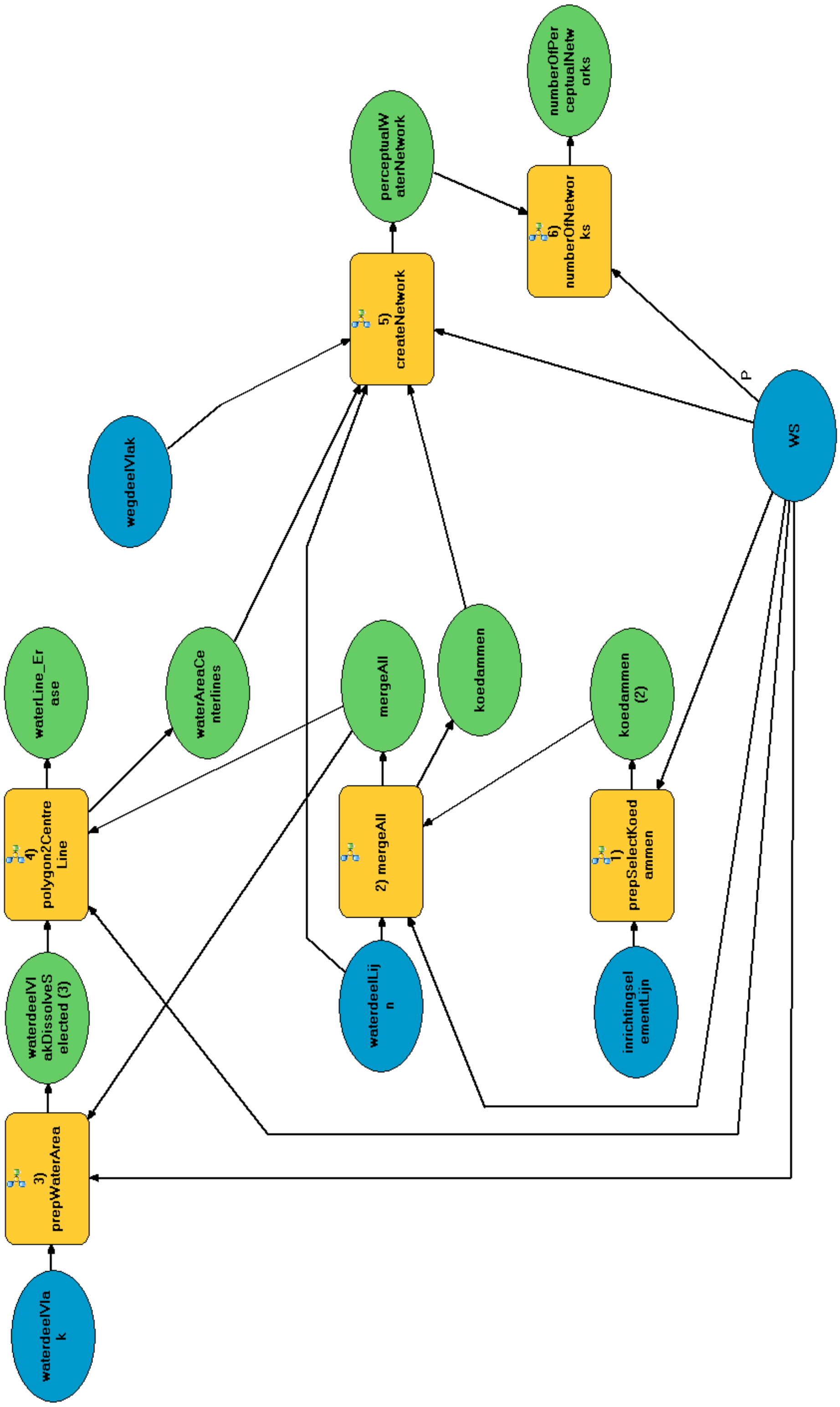


Figure 70: main model to improve network connectivity and reduce number of networks

The algorithm starts with data directly taken from Top10NL: water areas (*waterdeelVlak*), waterlines (*waterdeelLijn*) and a subclass of specific terrain constructions (*inrichtingselementLijn*). This data is topologically checked and the algorithm uses them to create an improved network. This is accomplished in six sub models.

The first sub model *1) prepSelectCowDykes* selects all cow dykes (line features) from specific terrain construction (*inrichtingselementLijn*) and creates a new feature class.

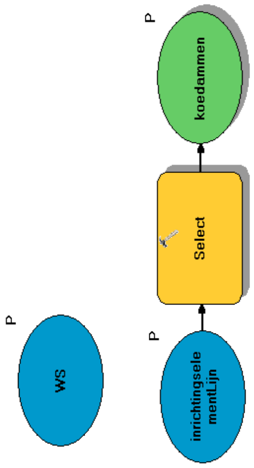


Figure 71: prepare select cow dykes

The second sub model *2) mergeAll* snaps over- and undershoots of cow dykes to waterlines (*waterdeelLijn*) to ensure the topological validity of the network. Subsequently the cow dykes and waterlines are merged into a new dataset, mergeAll.

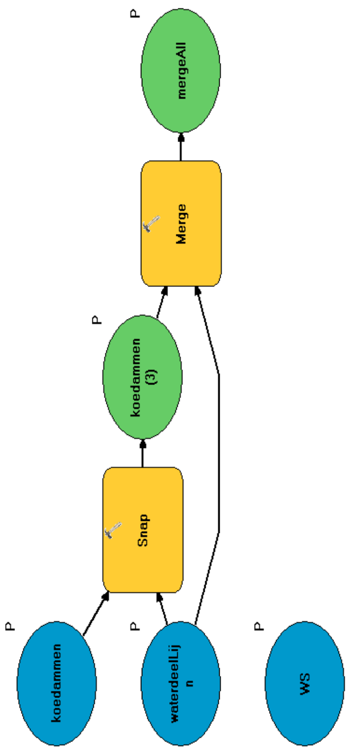


Figure 72: merge all

The third and fourth sub model deal with water polygons. The third sub model *3) prepareWaterArea*, detects isolated waterAreas (which are not touched by other water polygons or by waterlines or cow dykes) and removes them from the selection. All other water areas are part of a larger network and should be considered in network analysis. These are copied.

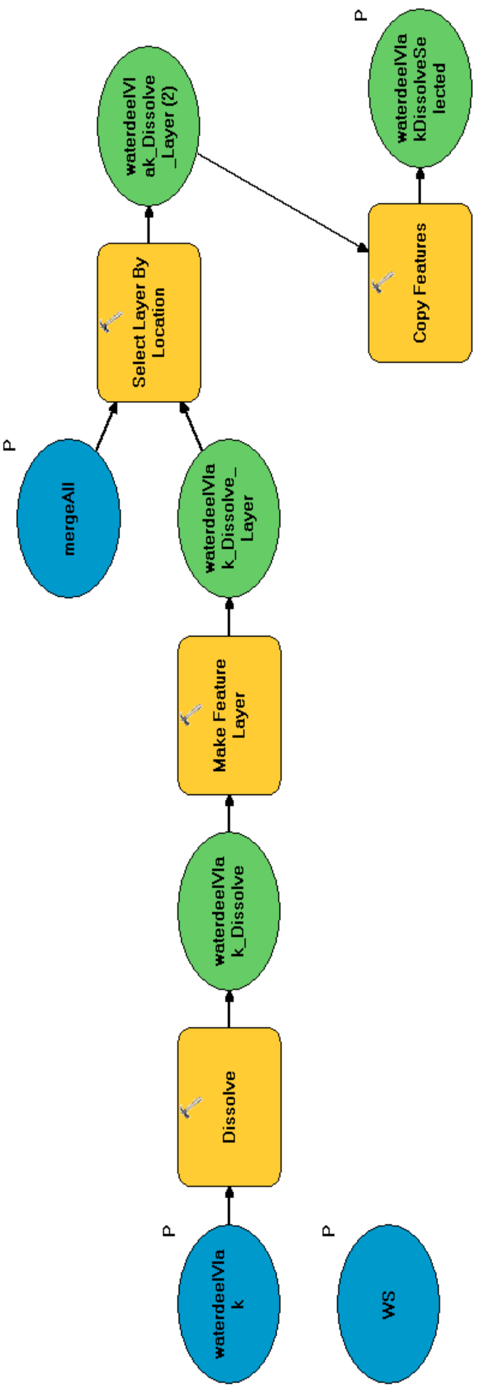


Figure 73: prepare water area

The fourth sub model *4) polygon2Centreline* replaces non-isolated water polygons with their calculated centrelines. This model was developed by my colleague Jan Bakermans. Some slight adaptations have been made.

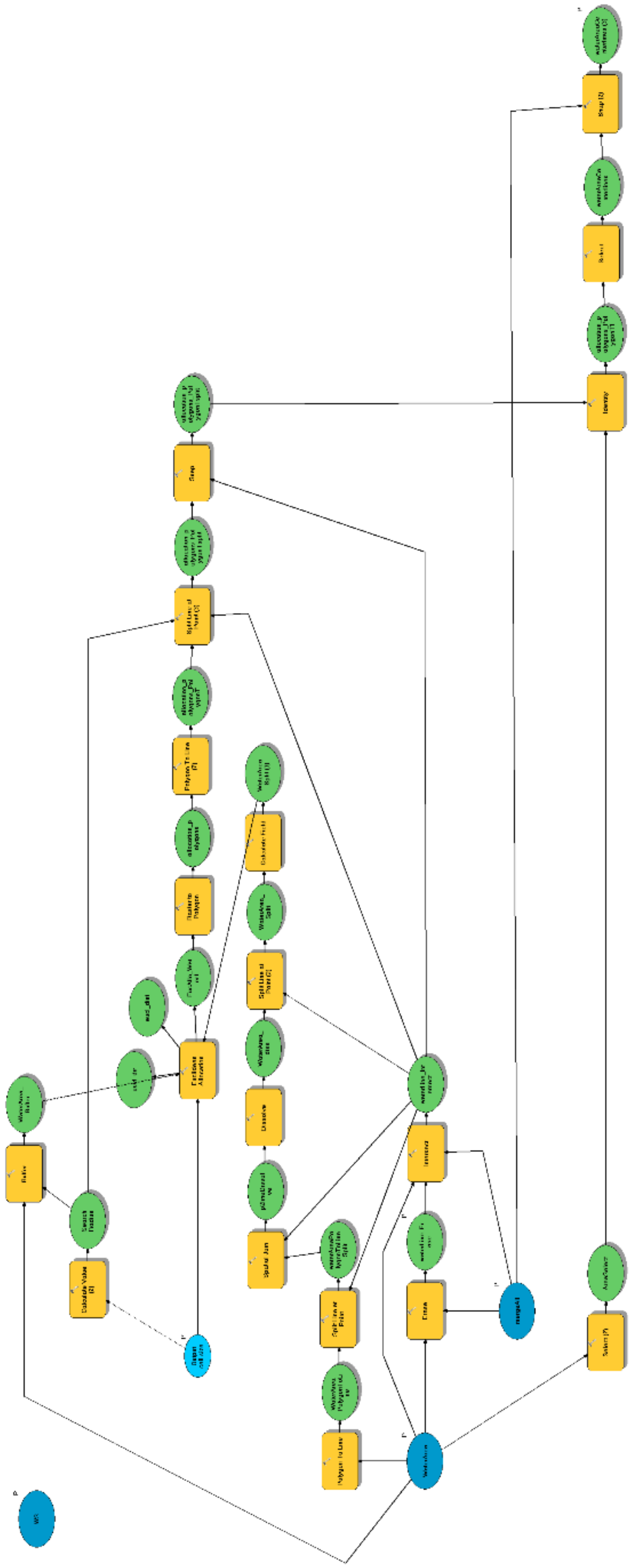


Figure 74: polygon to centreline

In the fifth sub model 5*) createNetwork*, the set of merged waterlines and cow dykes is supplemented with the computed centrelines. In addition, artificial perceptual links are also computed and added to the dataset.

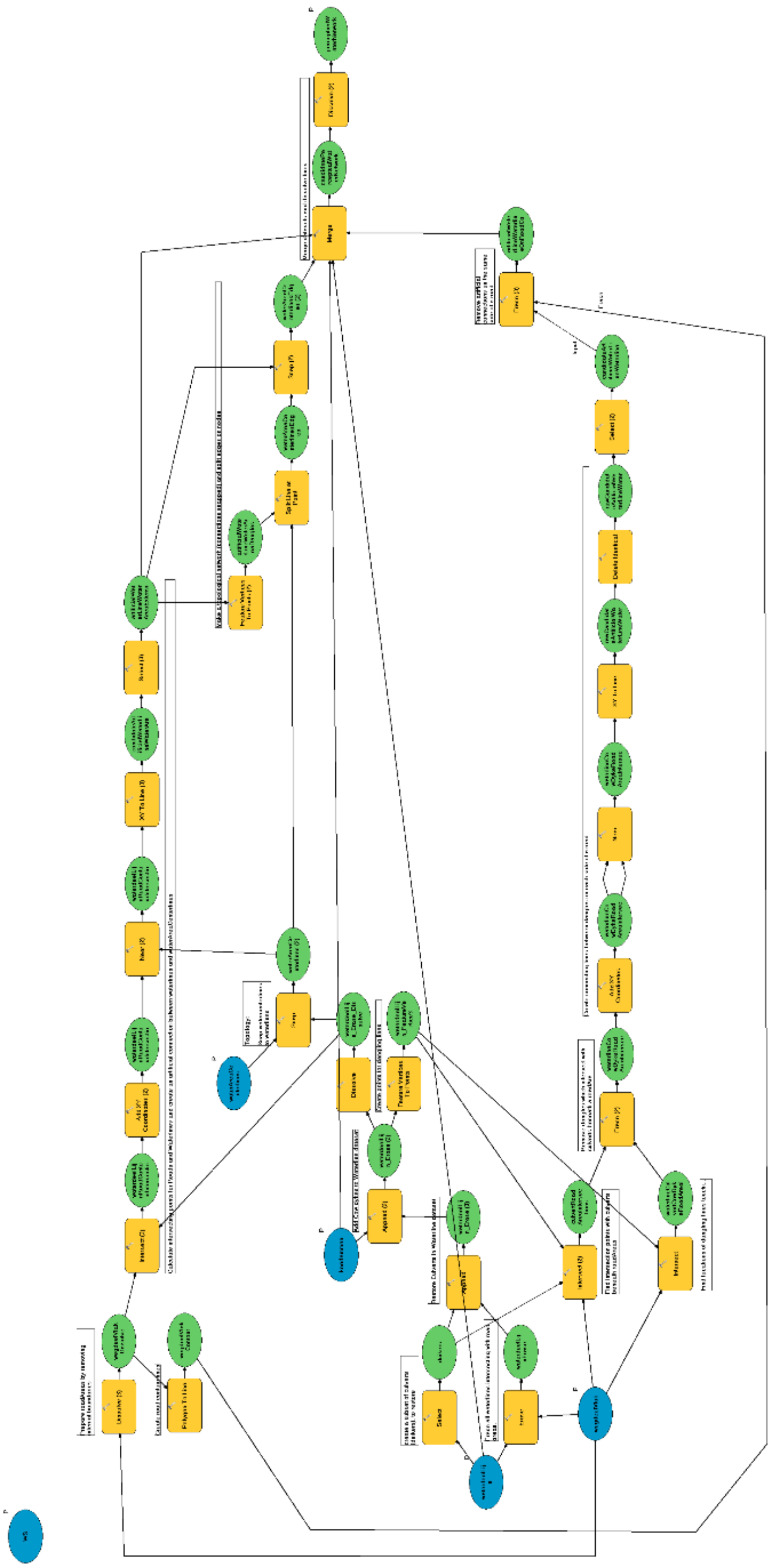


Figure 75: create network

Appendix G. Thin Road network

Preparation: Copy input feature class to secure input data from being edited. Add hierarchy field for processing purposes. Add invisibility field for processing purposes. Set value to 1 (all geometries the same hierarchy value). Thin Network with a minimum length of 300 meters.

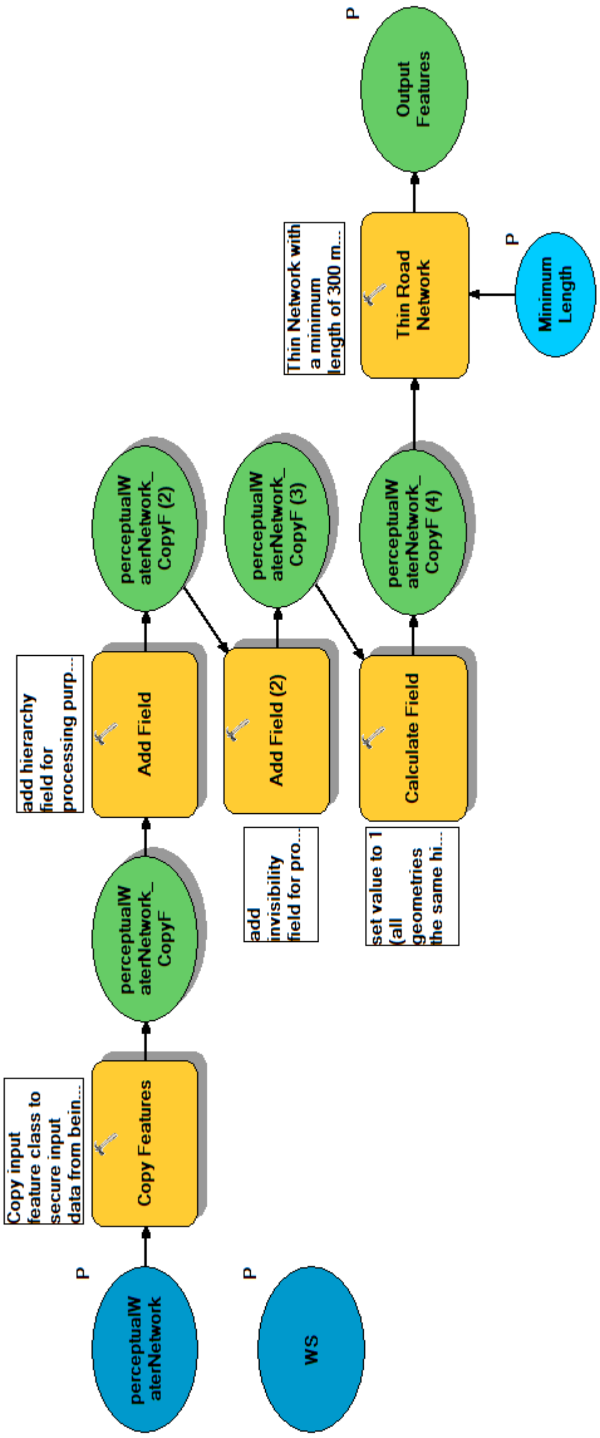


Figure 76: simple model to thin network

Appendix H. Stroke-based algorithm

Create new feature class to secure input data from being altered. Add and calculate field to save original ObjectID. Create Topological line structure by breaking lines at vertices and rebuilding the line features using Unsplit. Create point features for on start and end points of lines. Remove identical points and points on dangling lines.

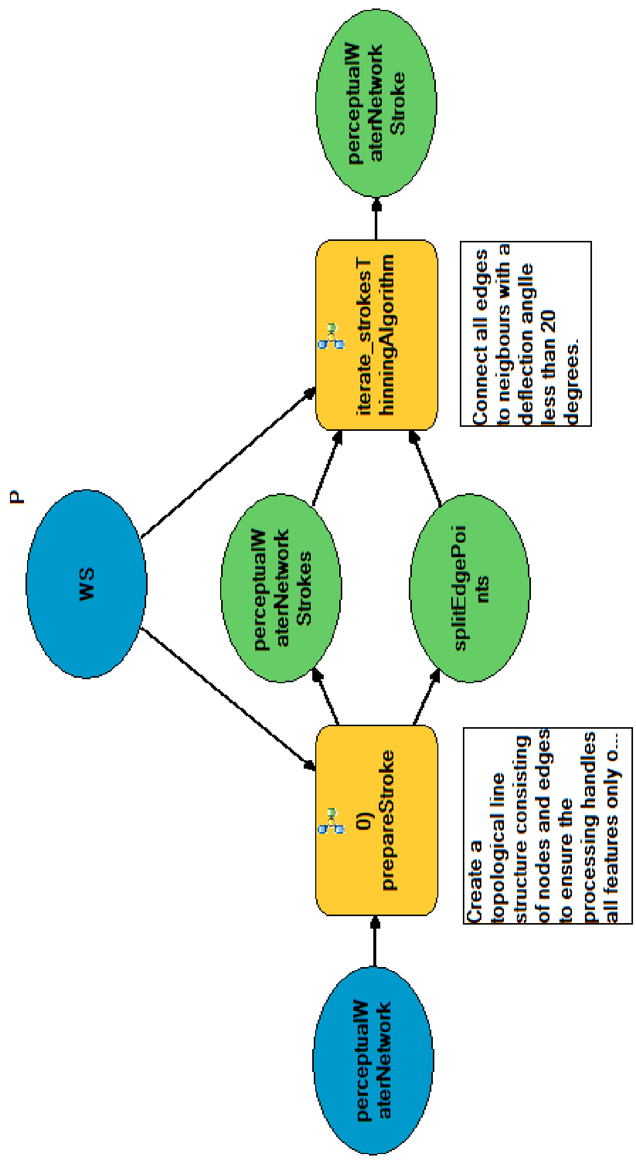


Figure 77: main model for stroke-based approach

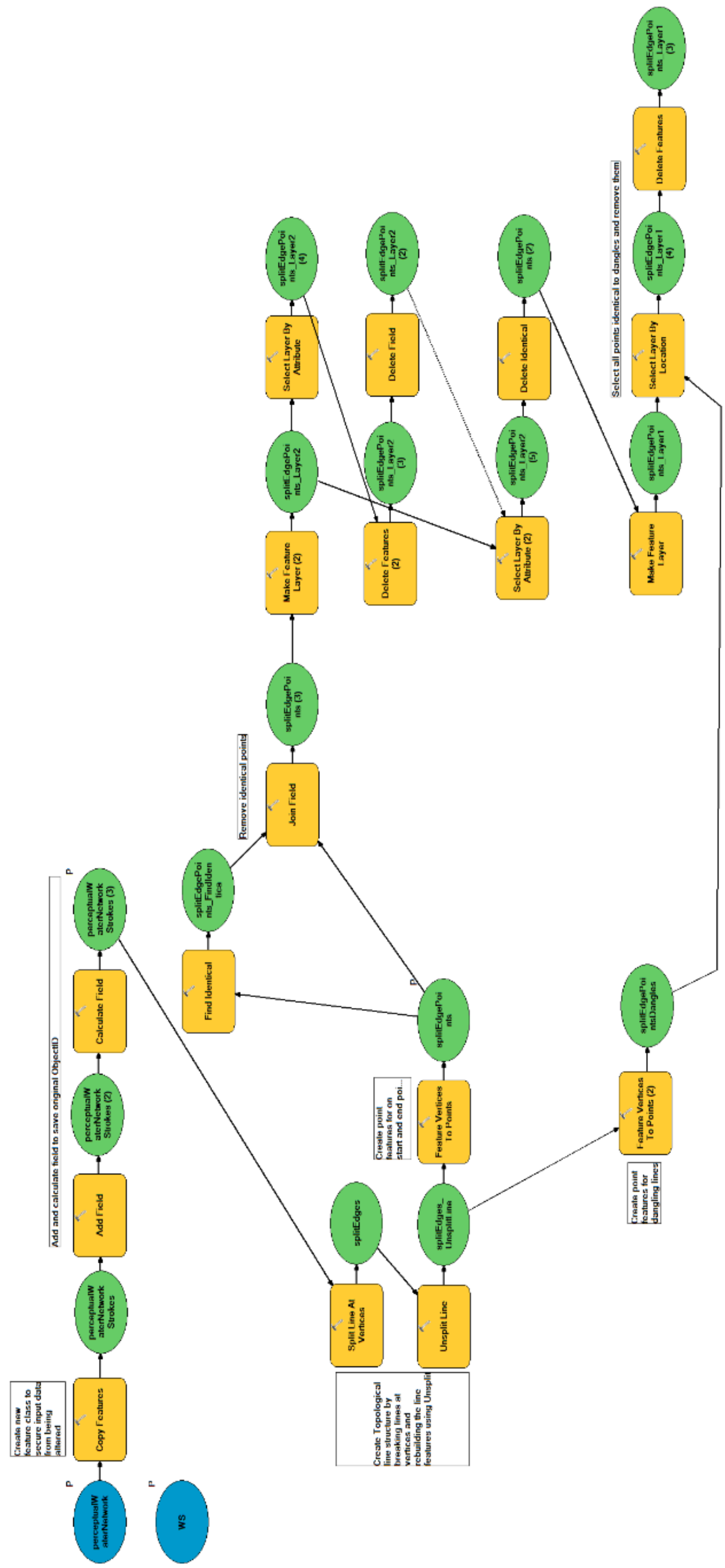


Figure 78: sub model to prepare topological structure with edges and nodes

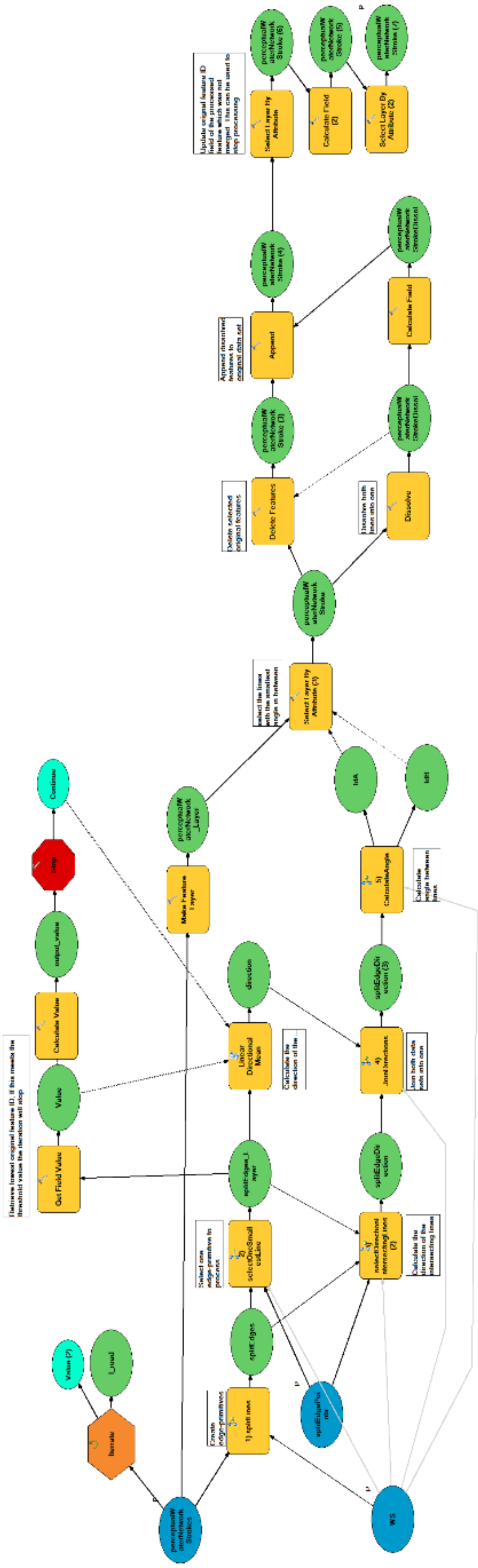


Figure 79: iterate strokes thinning algorithm

Create edge-primitives. Select one edge-primitive to process. Calculate the direction of the intersecting lines. Calculate the direction of the "first edge". Join both datasets into one. Calculate angle between lines. Select the lines with the smallest angle in between. Dissolve both lines into one and update original feature ID field of the processed feature which was merged. Delete selected original features. Append dissolved features to original dataset. Update original feature ID field of the processed feature which was not merged. This can be used to stop processing.

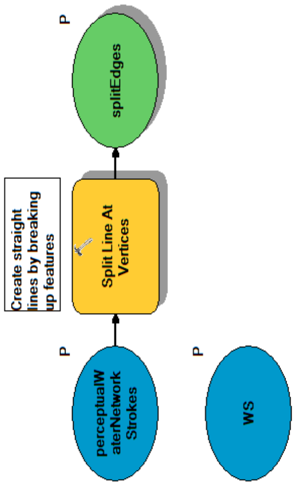


Figure 80: first sub model to create edge primitives to calculate directions with

Create edge primitives by breaking up features

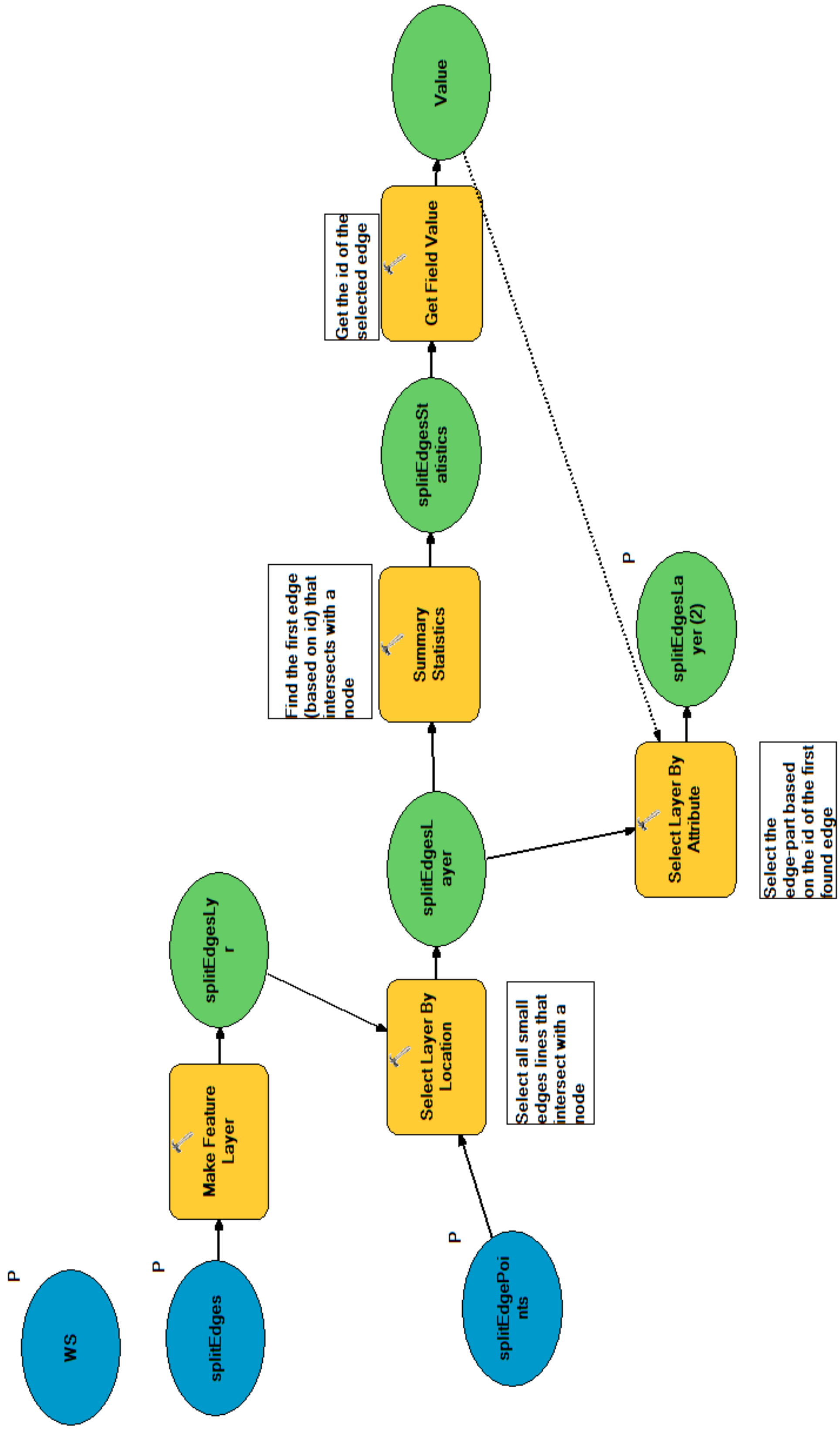


Figure 81: sub model to select one edge primitive which intersects with a node

Select all small edges lines that intersect with a node. Find the first edge (based on id) that intersects with a node. Get the id of the selected edge. Select the edge-part based on the id of the first found edge

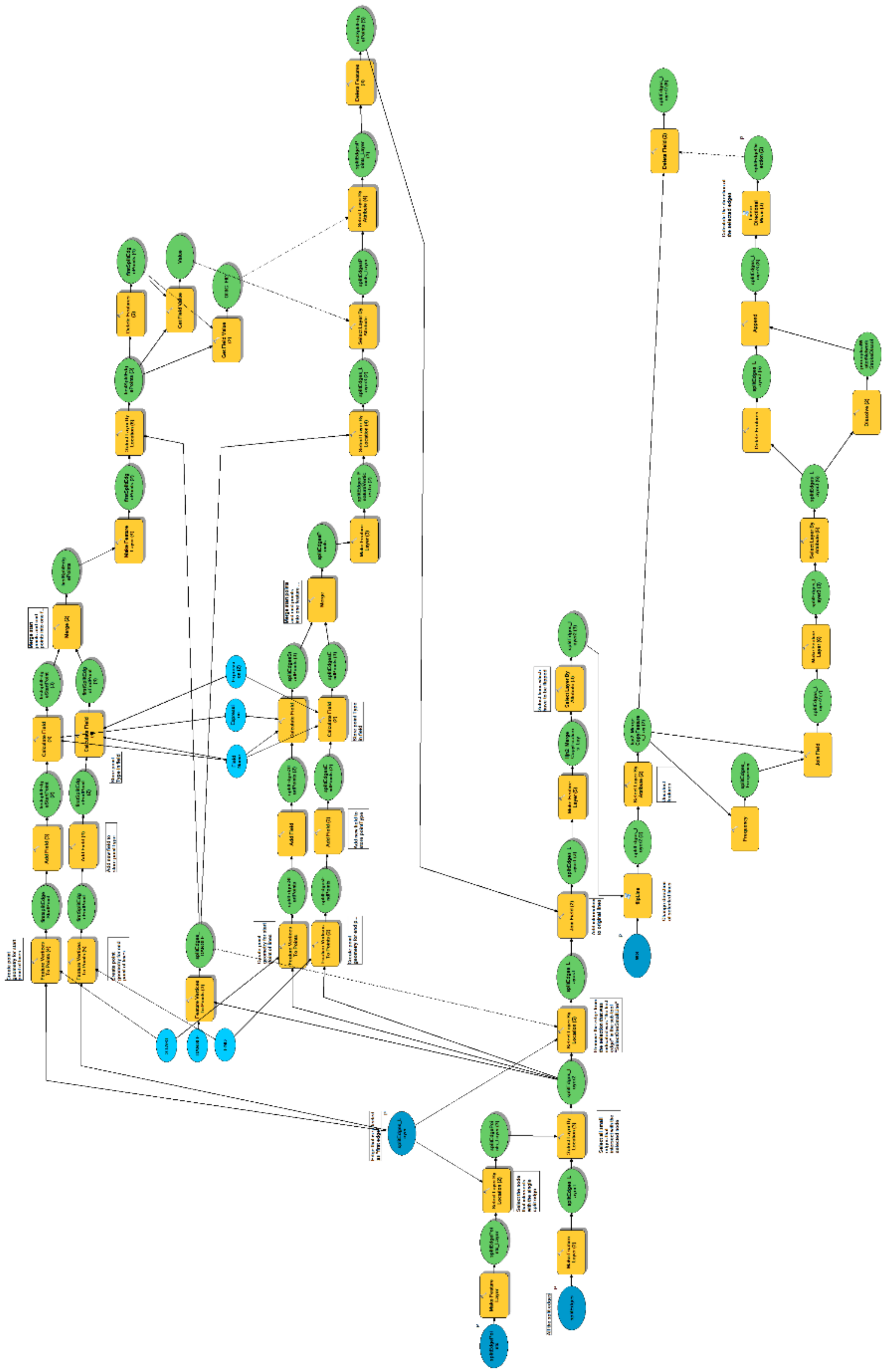


Figure 82: select direction of intersecting lines

Select the node that intersects with the selected edge primitive. Select all edge primitives that intersect with the selected node. Remove the edge primitive from the selection that was indicated was "the first edge" in the sub tool "SelectOneSmallLine". Establish the orientation of all selected primitives and compare these with the orientation of the first edge primitive. All edge primitives which orientation are not in line with the first edge's orientation are flipped (by the python script below) and for each edge primitive the direction is computed.

```
#-------------------------------------------------------------------------------
# Name:        Flip line
# Purpose:     Change direction of input features. This can be a feature layer
#              or a feature class
#
# Author:      altenv
#
# Created:     31-10-2014
# Copyright:   (c) altenv 2014
# Licence:     <your licence>
#-------------------------------------------------------------------------------

#set env
import arcpy
from arcpy import env

env.workspace = arcpy.GetParameterAsText(0)
inFeatures      = arcpy.GetParameterAsText(1)

# operation
try:
   arcpy.FlipLine_edit(inFeatures)

# error handling
except Exception, e:
   # If an error occurred, print line number and error message
   import traceback, sys
   tb = sys.exc_info()[2]
   print "Line %i" % tb.tb_lineno
   print e.message
```

Figure 83: Python script to change the direction of a line

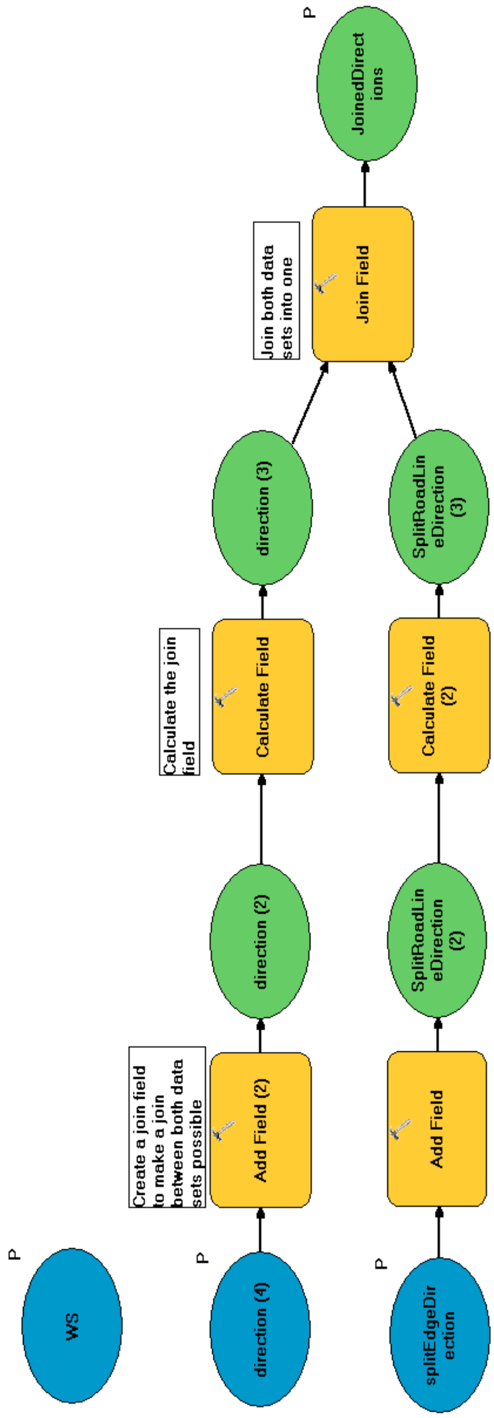


Figure 84: sub model to join directions

Create a join field to make a join between both datasets possible. Calculate the join field. Join both datasets into one.

Add angle field. Calculate the deflection angle. Select deflection angles smaller than 20 degrees. Find the first edge (based on id) that intersects with a node. Get the id of the selected edge. Select the edge based on the id of the edge with the smallest deflection angle. Get id of both edges.

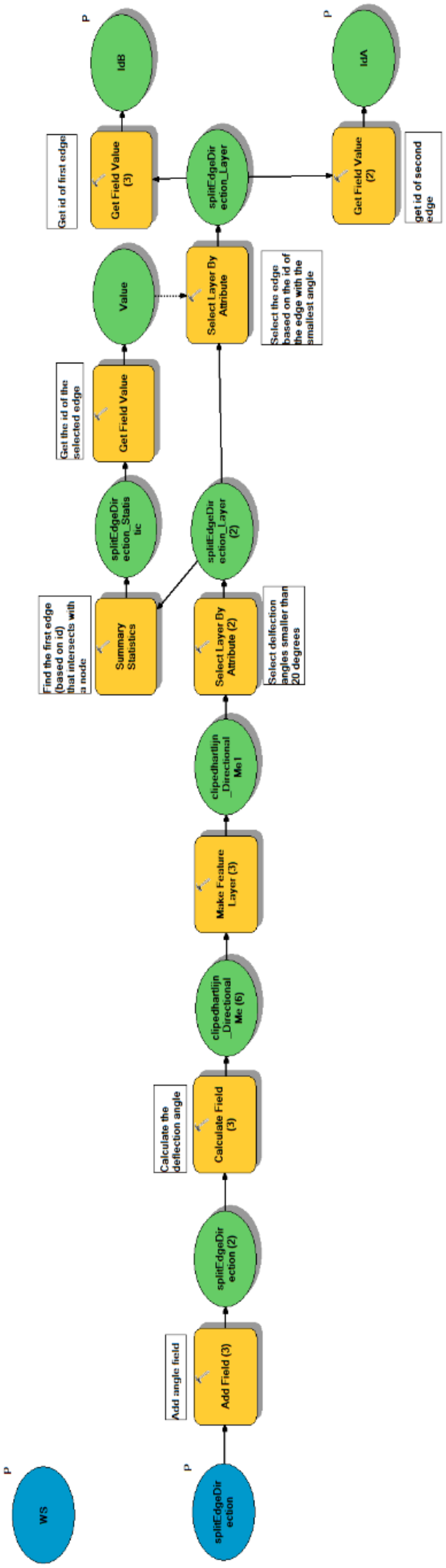


Figure 85: sub model to calculate deflection angle of lines

Appendix I. Mesh-based algorithm

Before starting the iteration, a preparatory model is run once which performs three actions to validate the data: 1) creation of an outer boundary based on the feature envelope. This is necessary to prevent the emerging of isolated meshes. Isolated meshes cannot be added to a neighbouring mesh, and will cause an endless loop; 2) removal of slivers from the dataset by integrating lines. Lines will be collapsed if they are within 1 meter of each other; 3) removal of overlapping features which are caused by the previous integration. First, all lines are split by their vertices, next duplicate lines are removed and last lines are unsplit, thus removing pseudo nodes.

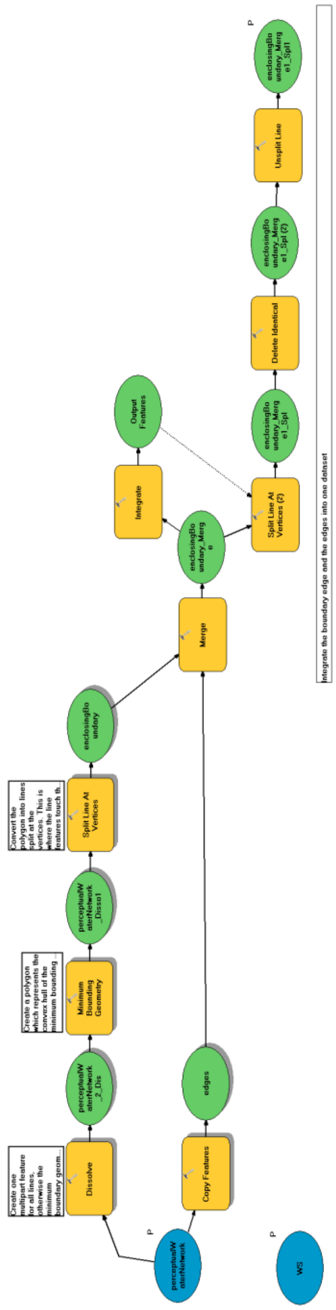


Figure 86: preparatory model

After this preparation the meshes can be processed. The algorithm used is based on the work of (Post 2014) for his master thesis. The model has been adapted on two points: the FME algorithm to select the densest mesh has been rebuilt in ArcGIS ModelBuilder. This is more convenient and speeds up the process, since FME does not have to be launched for each iteration. Some alterations in calculating the threshold and Density values have been made to get the iteration working.

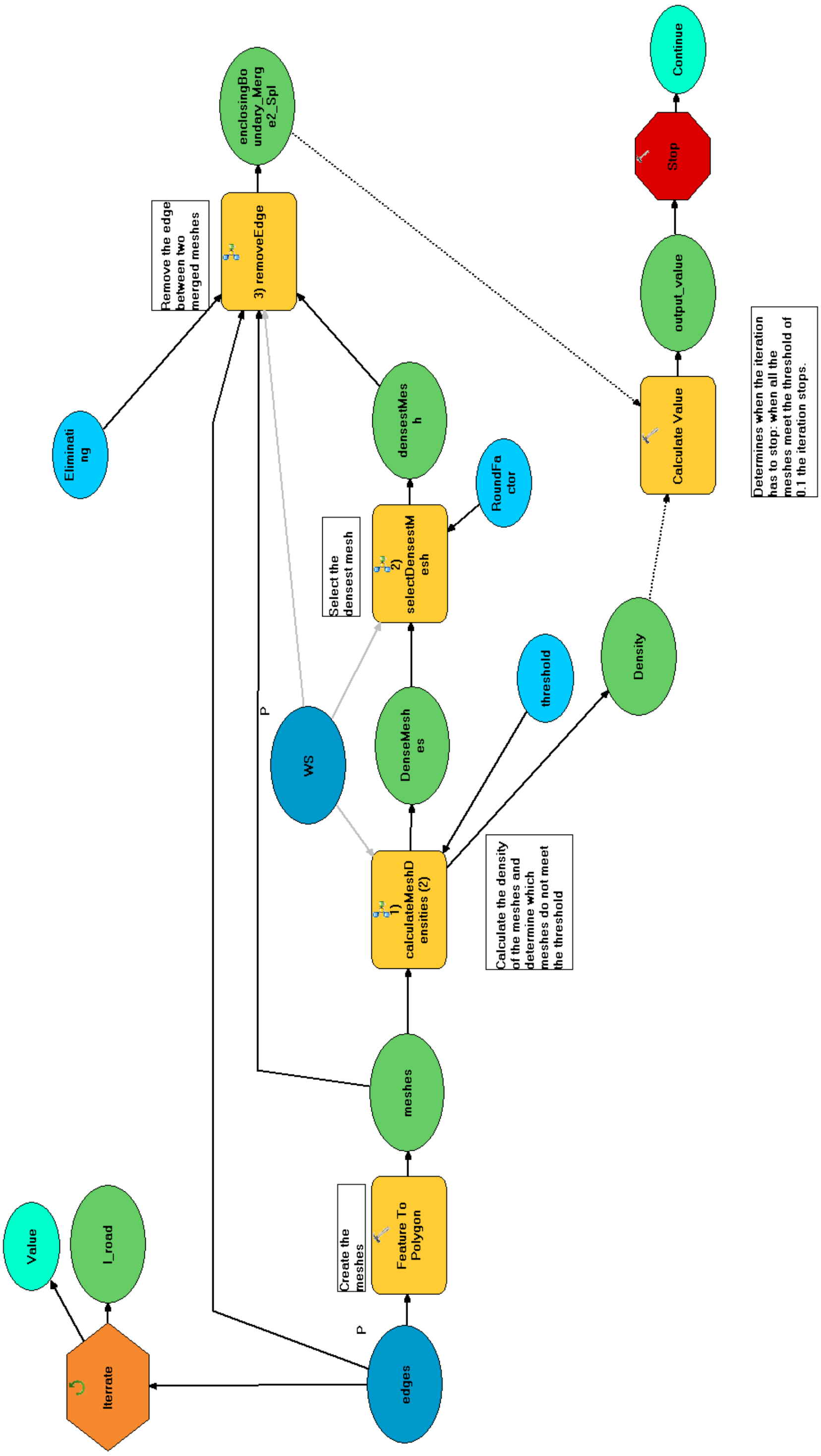


Figure 87: mesh-based algorithm

The first sub model *1) calculateMeshDensities:* calculates the density of the meshes and determines which meshes do not meet the threshold.

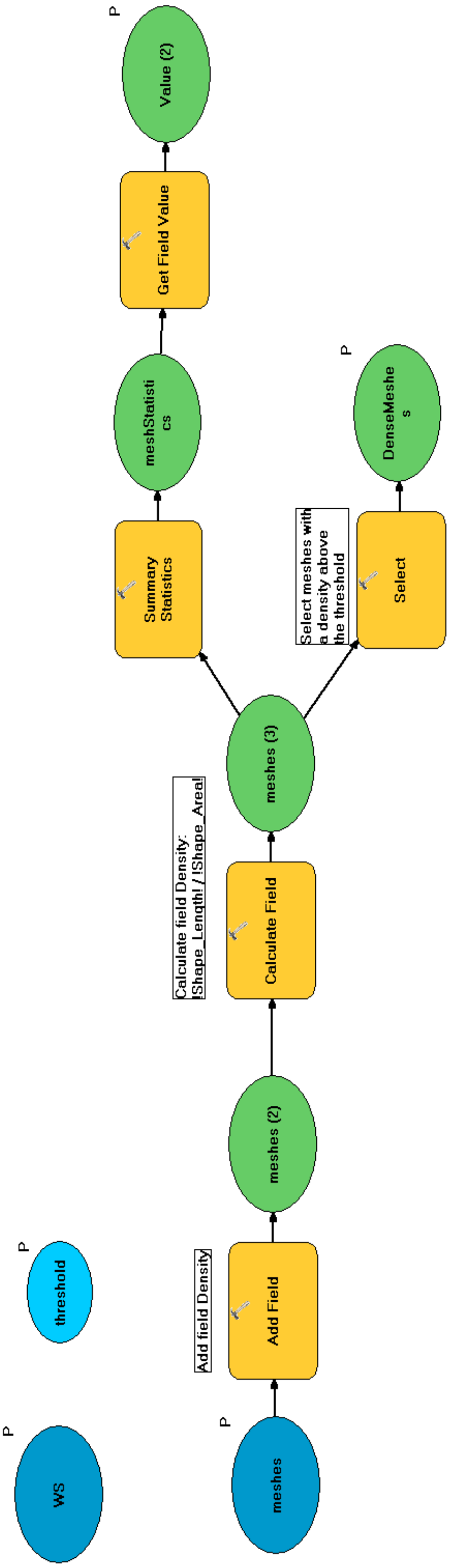


Figure 88: sub model to calculate mesh densities

This is done by first creating a field Density which then is populated by a computation for each mesh of the density. Meshes with a density above the threshold value are copied to a separate feature class for processing purposes. By applying summary statistics the mesh with the highest Density value is selected and this value is output for use in subsequent models.

The second sub model *2) selectDensestMesh* selects the densest mesh in the set of meshes.

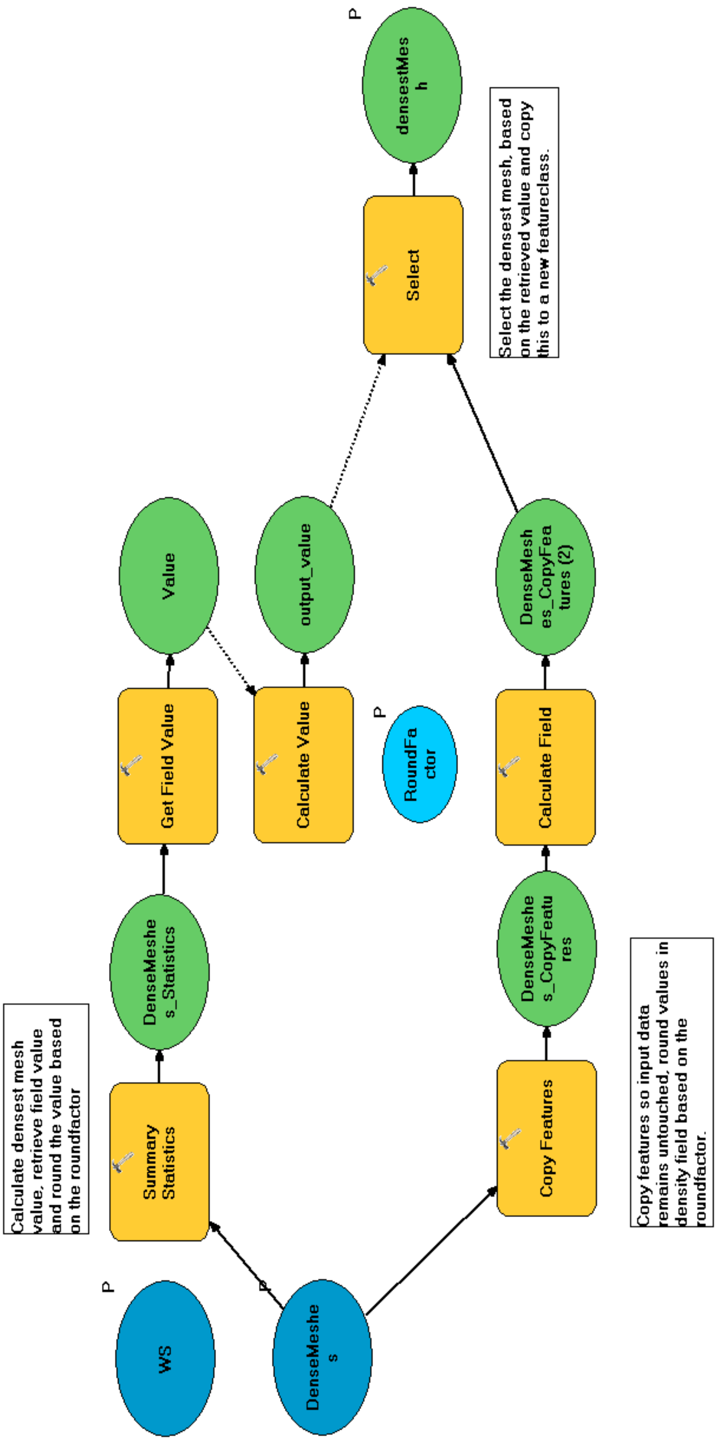


Figure 89: sub model to select the densest mesh

This is done by calculating the densest mesh value using summary statistics. (To keep the model valid, this value is rounded based on a predefined round factor). Input features are copied so input data remains untouched. Round values in density field based on the round factor. Select the densest mesh, based on the retrieved value and copy this to a new feature class.

The third sub model *3) removeBorderLine* removes the edge between two merged meshes.

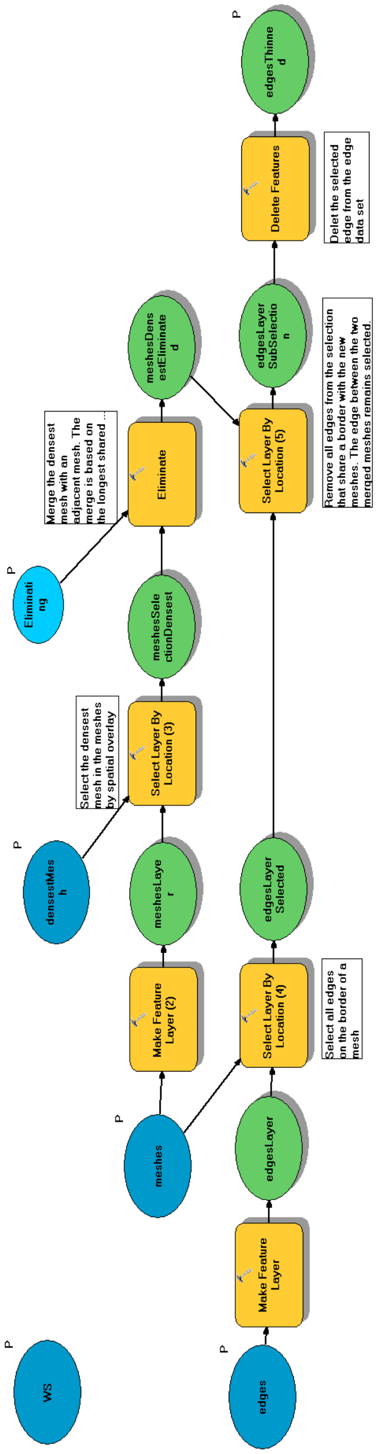


Figure 90: sub model to remove edge between two merged meshes

It first selects all edges on the border of a mesh. Next, it selects the densest mesh in the meshes by spatial overlaying and merges this mesh with an adjacent mesh. The merge is based on the longest shared border. From the initial selection all edges that share a border with the new meshes are removed. The edge between the two merged meshes remains selected. This edge is eliminated from the edge dataset.

Appendix J. Resemblance with input data

Resemblance with input data after thinning for partition 1

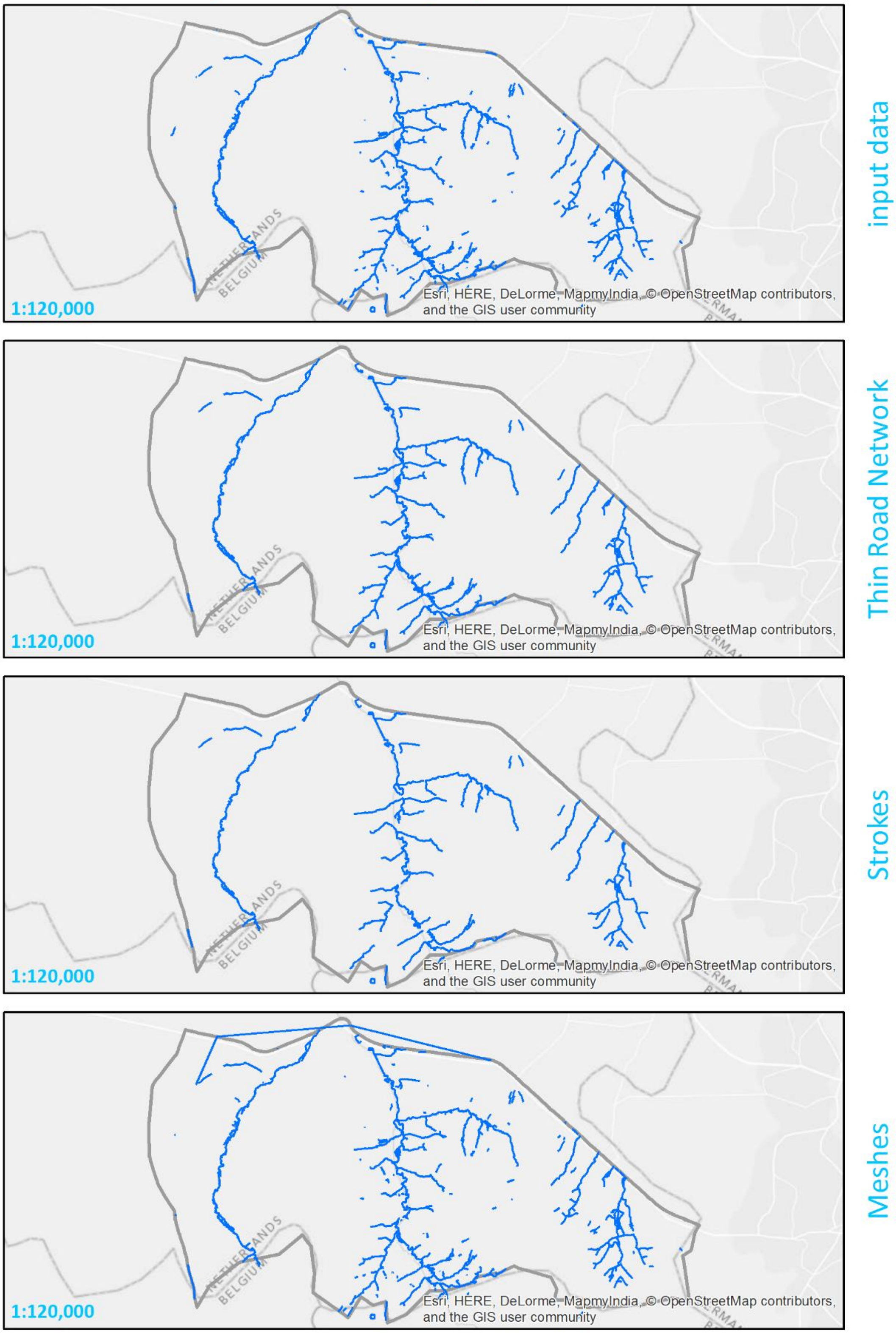


Figure 91: resemblance with input data after thinning for partition 1

Resemblance with input data after thinning for partition 17

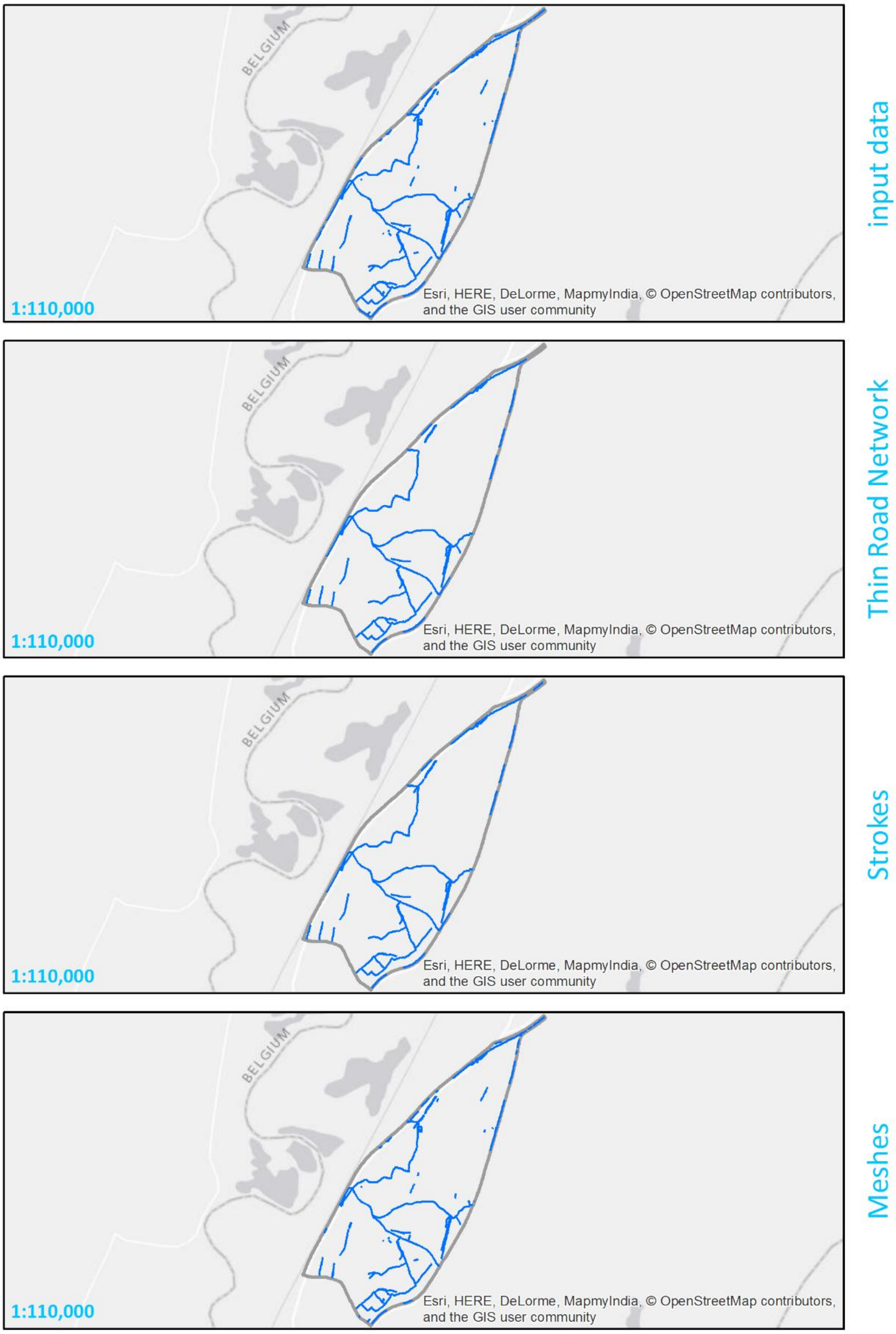


Figure 92: resemblance with input data after thinning for partition 17

Resemblance with input data after thinning for partition 49

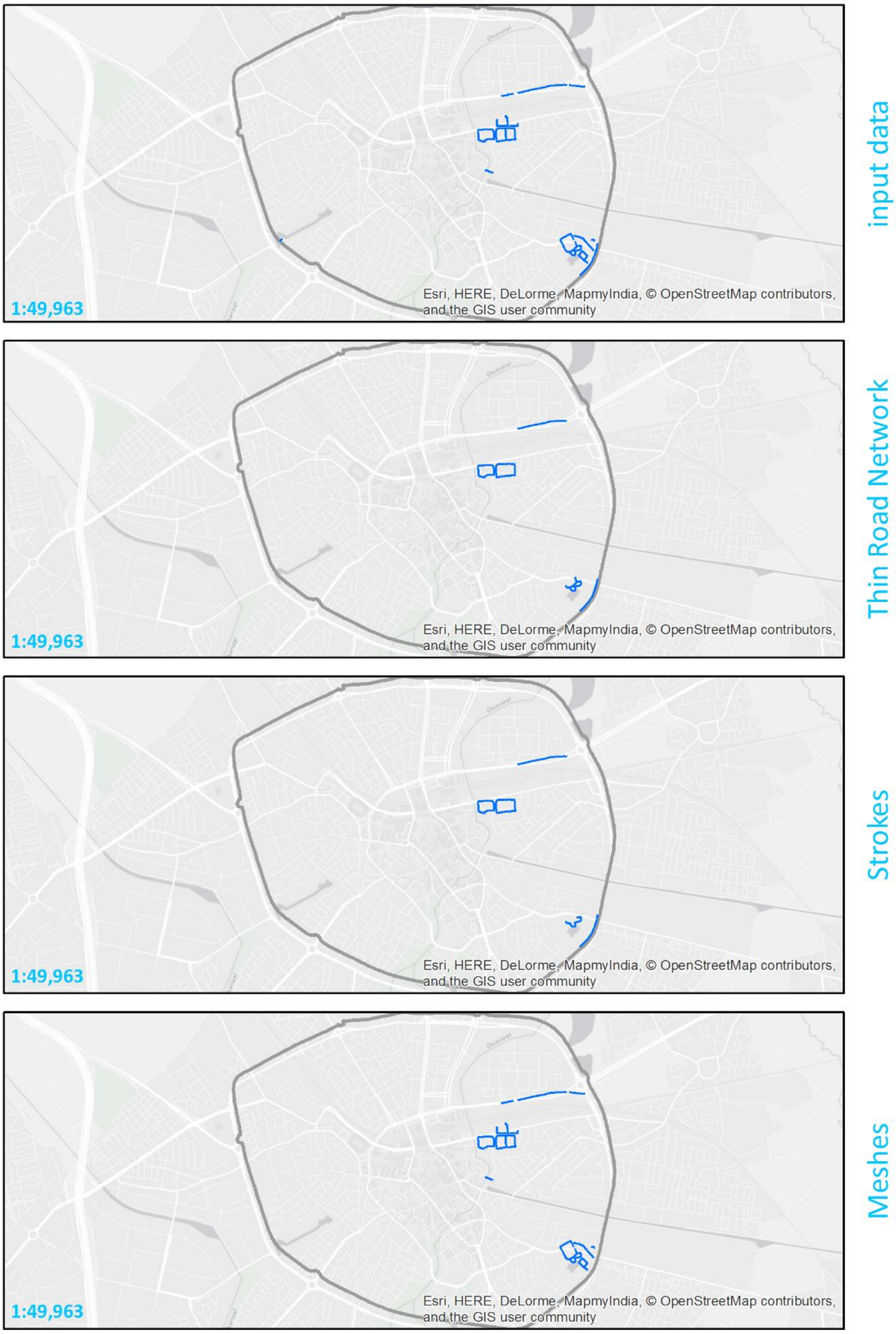


Figure 93: resemblance with input data after thinning for partition 49

Resemblance with input data after thinning for partition 93

Figure 94: resemblance with input data after thinning for partition 93

Resemblance with input data after thinning for partition 198

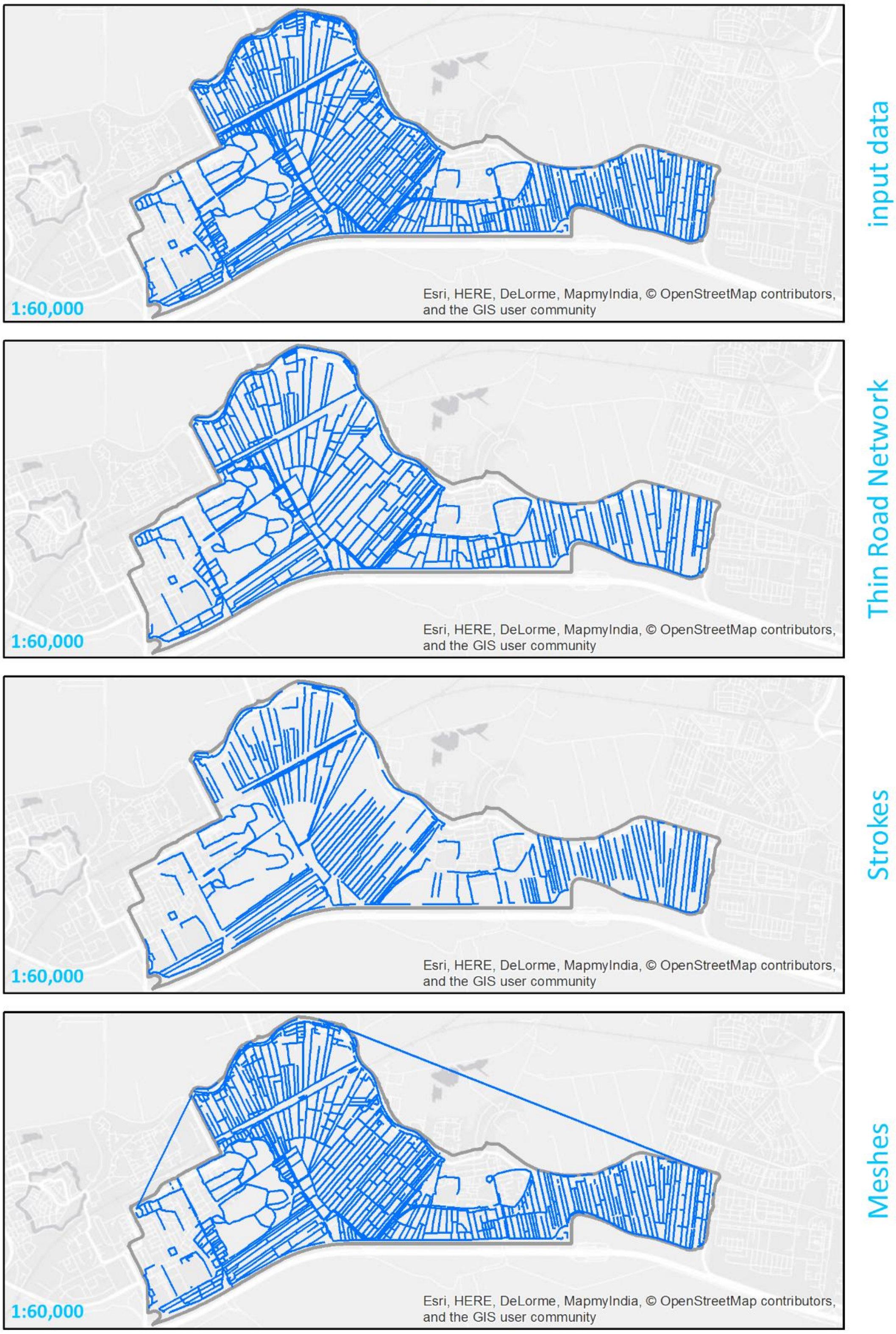


Figure 95: resemblance with input data after thinning for partition 198

Resemblance with input data after thinning for partition 217

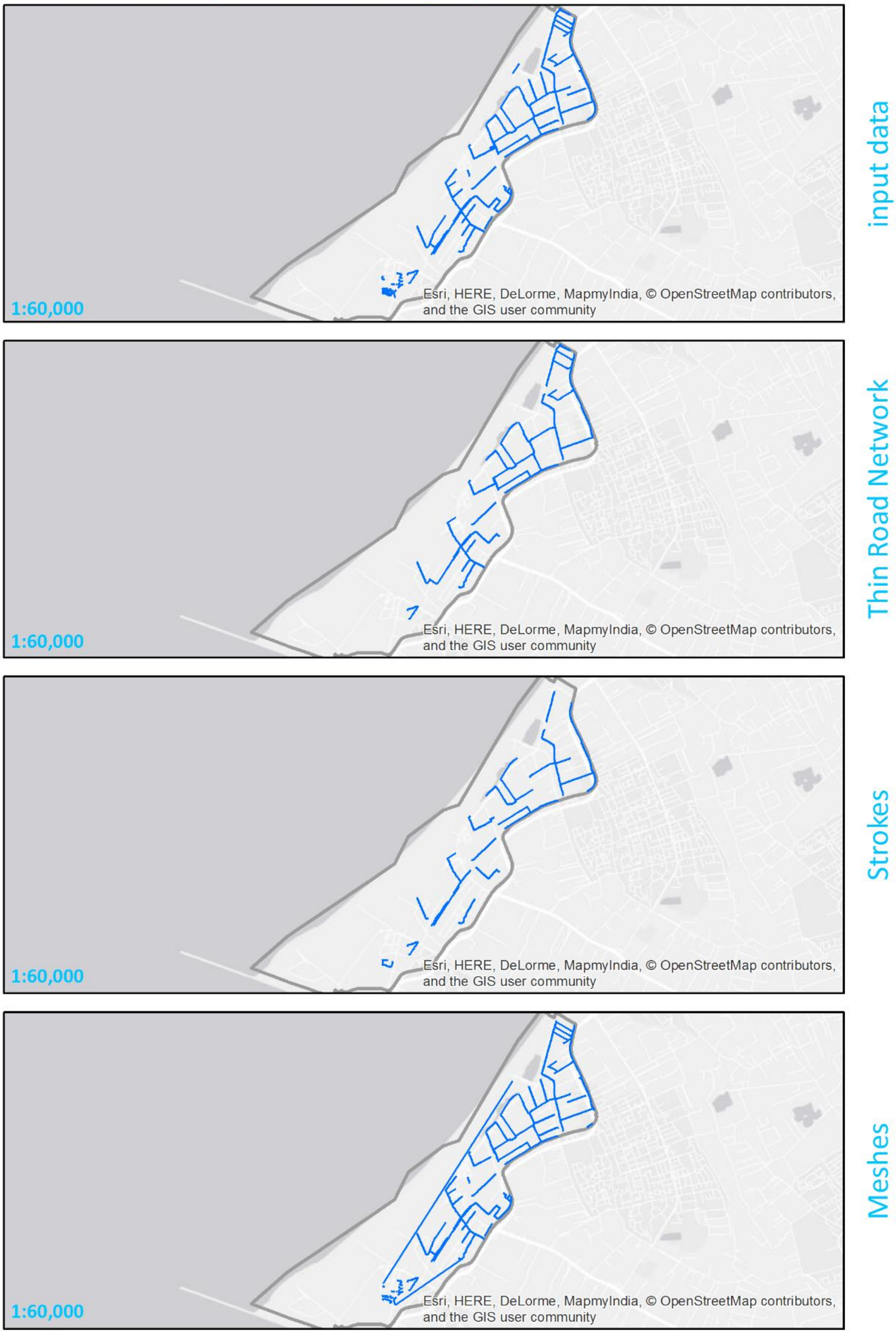


Figure 96: resemblance with input data after thinning for partition 217

Resemblance with input data after thinning for partition 260

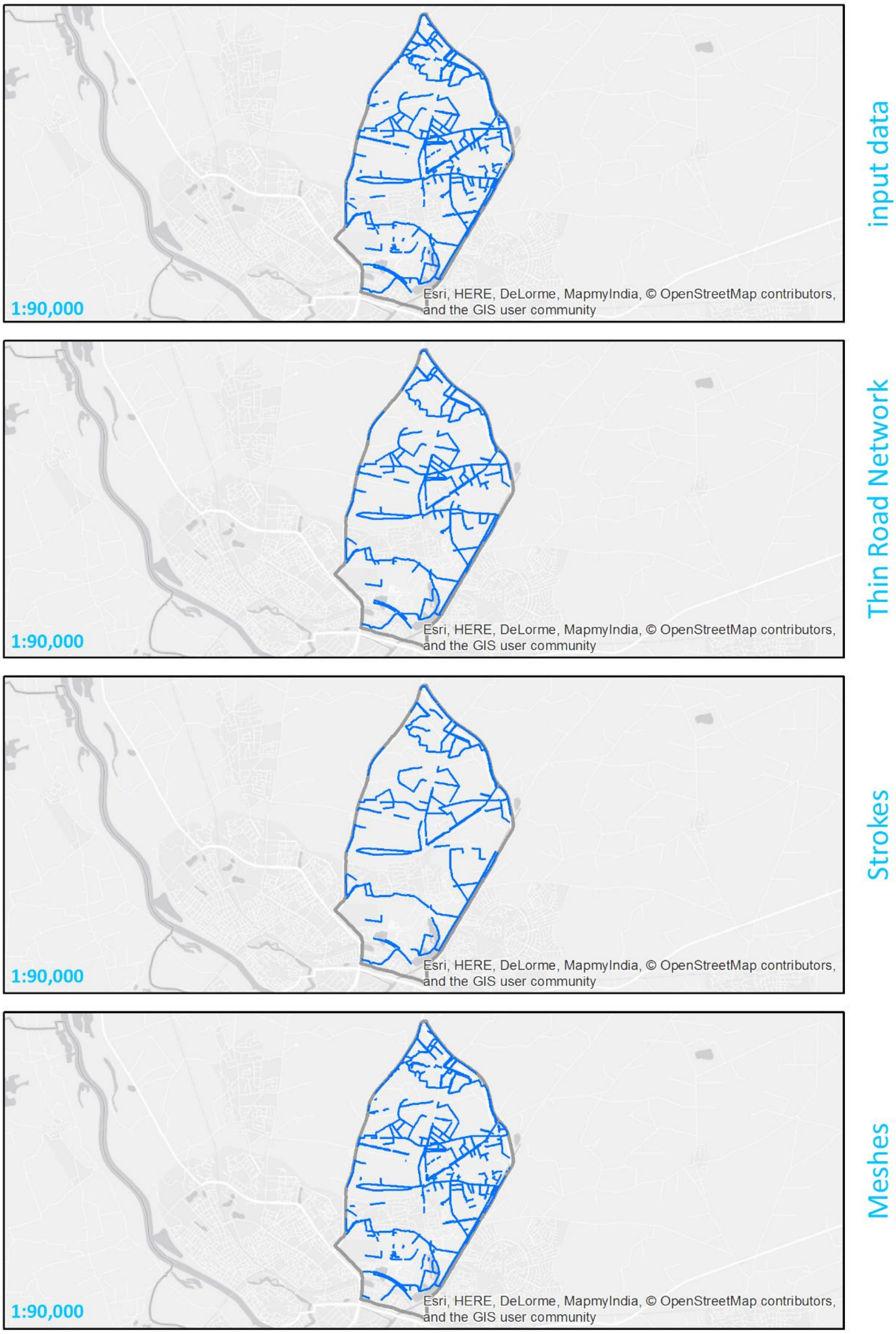


Figure 97: resemblance with input data after thinning for partition 260

Resemblance with input data after thinning for partition 298

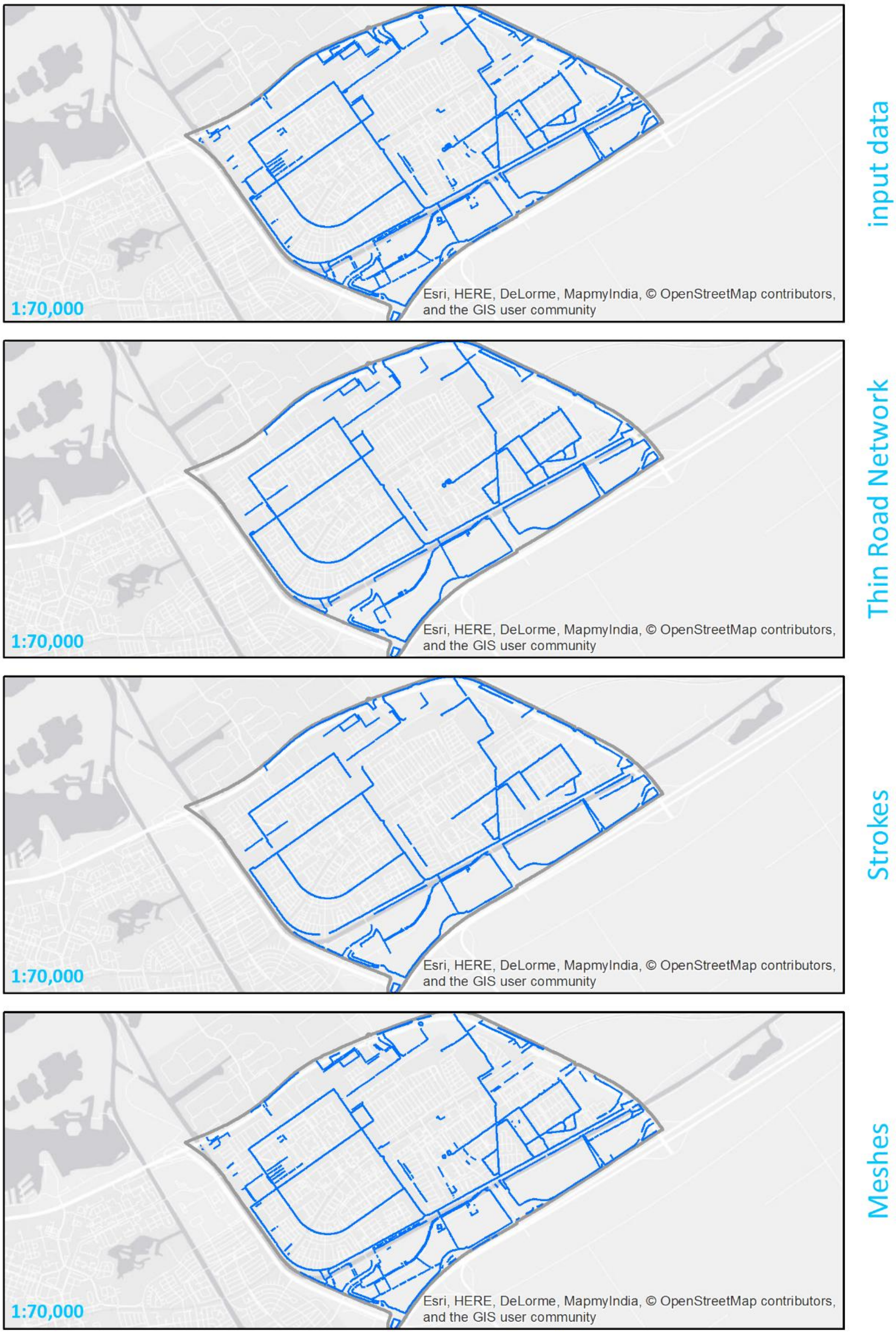


Figure 98: resemblance with input data after thinning for partition 298

Resemblance with input data after thinning for partition 345

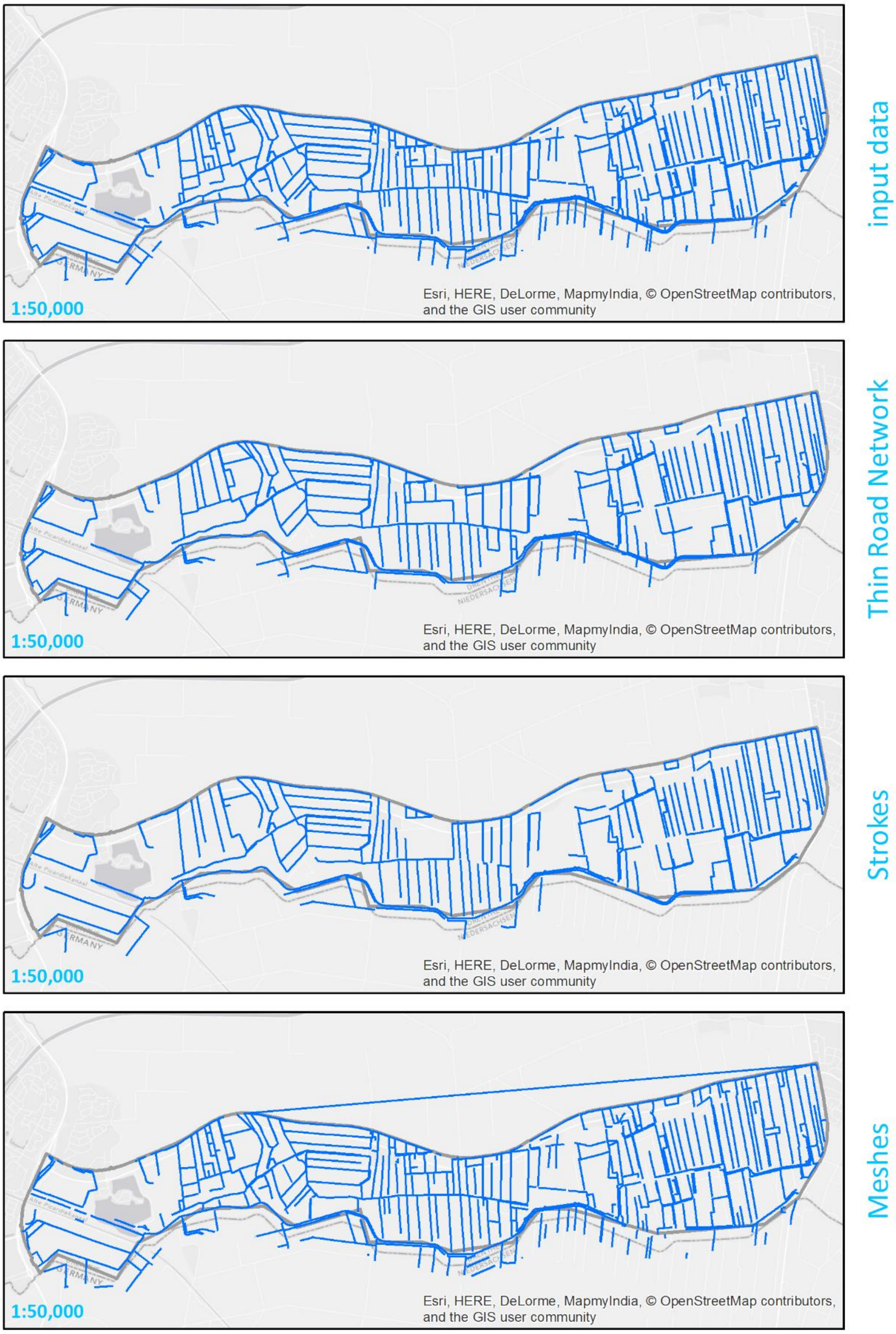


Figure 99: resemblance with input data after thinning for partition 345

Resemblance with input data after thinning for partition 403

Figure 100: resemblance with input data after thinning for partition 403

Resemblance with input data after thinning for partition 442

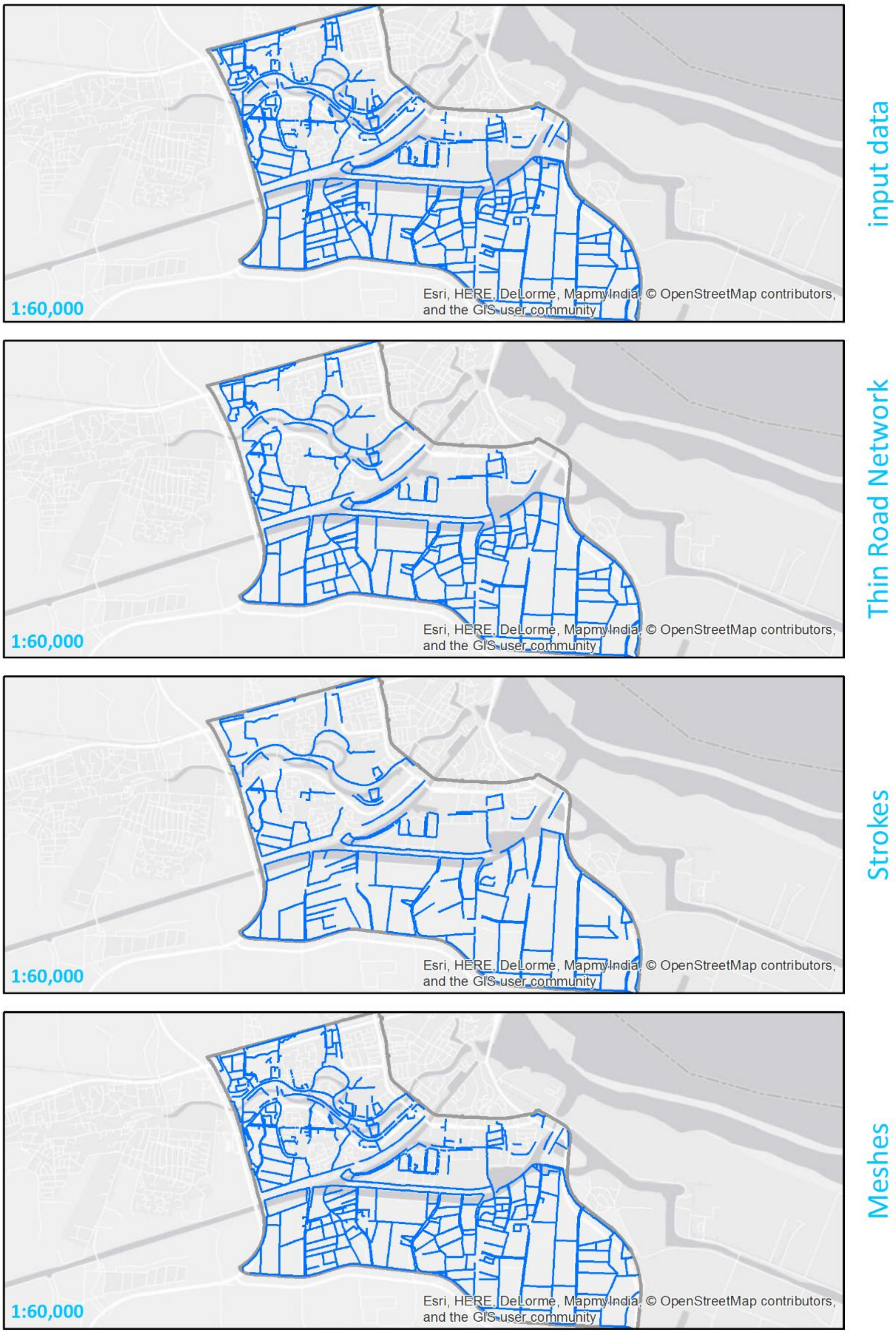


Figure 101: resemblance with input data after thinning for partition 442

Appendix K. Largest network analysis

This appendix contains two tables and eleven comparison maps. Table 13 displays data about the computed connectivity index for largest network per partition for each algorithm and Table 14 presents information about the circuit index for largest network per partition for each algorithm.

Table 13: connectivity index for largest network per partition for each algorithm

Partition	Largest network equal?	**γ – index input**	**γ – index TRN**	**γ – index strokes**	**γ – index meshes**
1	no	0.357	0.348	0.444	0.364
17	yes	0.328	0.333	0.344	0.312
49	yes	0.383	0.444	0.583	0.381
93	yes	0.361	0.396	0.388	0.343
198	no	0.375	0.369	0.406	0.369
217	no	0.337	0.327	0.381	0.333
260	yes	0.352	0.376	0.356	0.353
298	yes	0.401	0.408	0.327	0.262
345	yes	0.400	0.366	0.386	0.317
403	yes	0.371	0.360	0.317	0.284
442	no	0.344	0.401	0.411	0.346

Table 14: circuit index for largest network per partition for each algorithm

Partition	Largest network equal?	**α – index input**	**α – index TRN**	**α – index Strokes**	**α – index Meshes**
1	no	0.032	0.019	0.14286	0.130
17	yes	-0.011	-0.016	0.00000	0.000
49	yes	0.051	0.091	0.00000	0.285
93	yes	0.041	0.094	0.13600	0.076
198	no	0.063	0.053	0.00000	0.107
217	no	0.003	-0.019	0.00000	0.049
260	yes	0.027	0.061	0.06452	0.017

298	yes	0.097	0.104	0.02326	-0.019
345	yes	0.000	0.048	0.00000	0.072
403	yes	0.055	0.038	0.01124	-0.030
442	no	0.015	0.098	0.17073	0.113

Retained largest network after thinning for partition 1

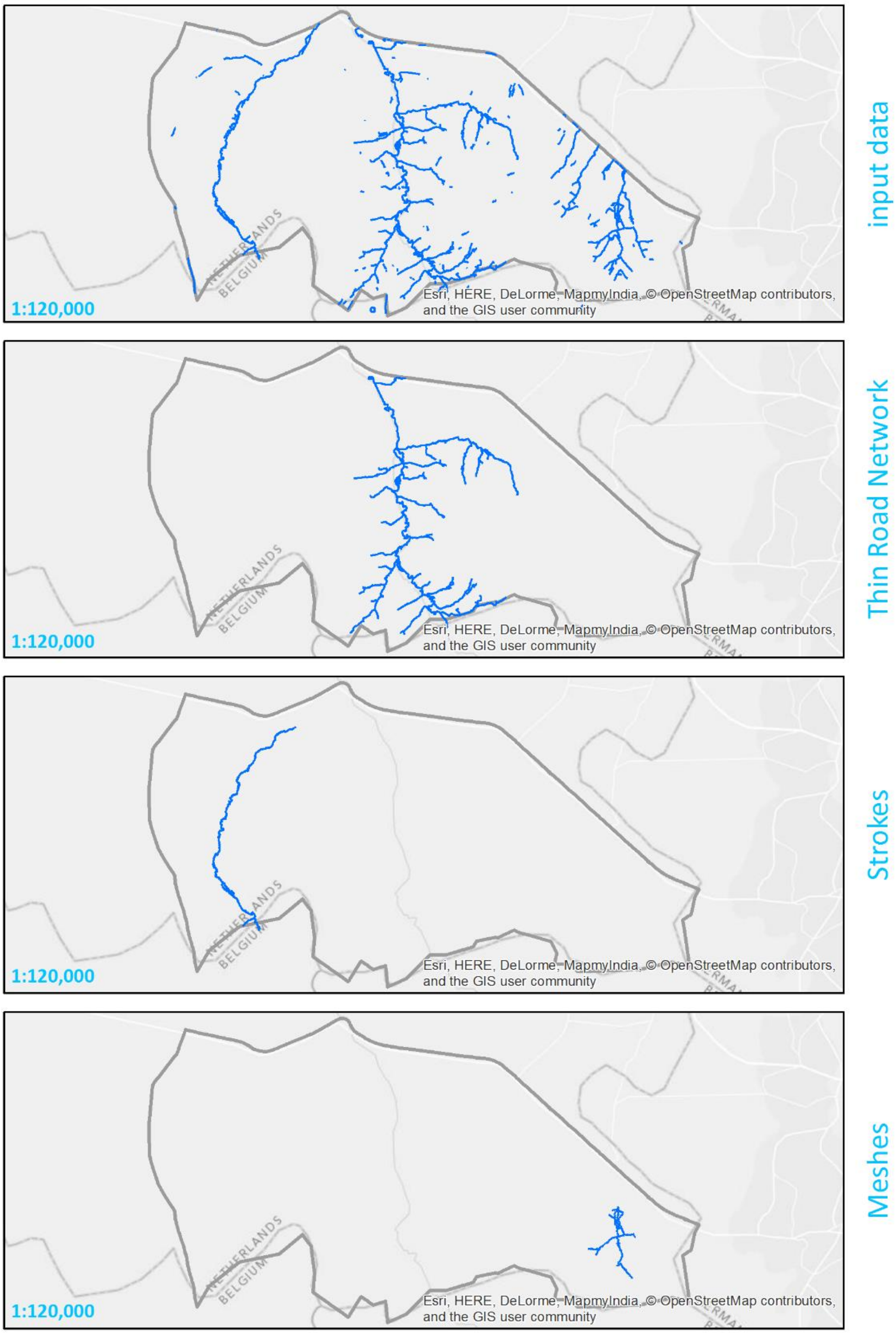


Figure 102: retained largest network after thinning for partition 1

Retained largest network after thinning for partition 17

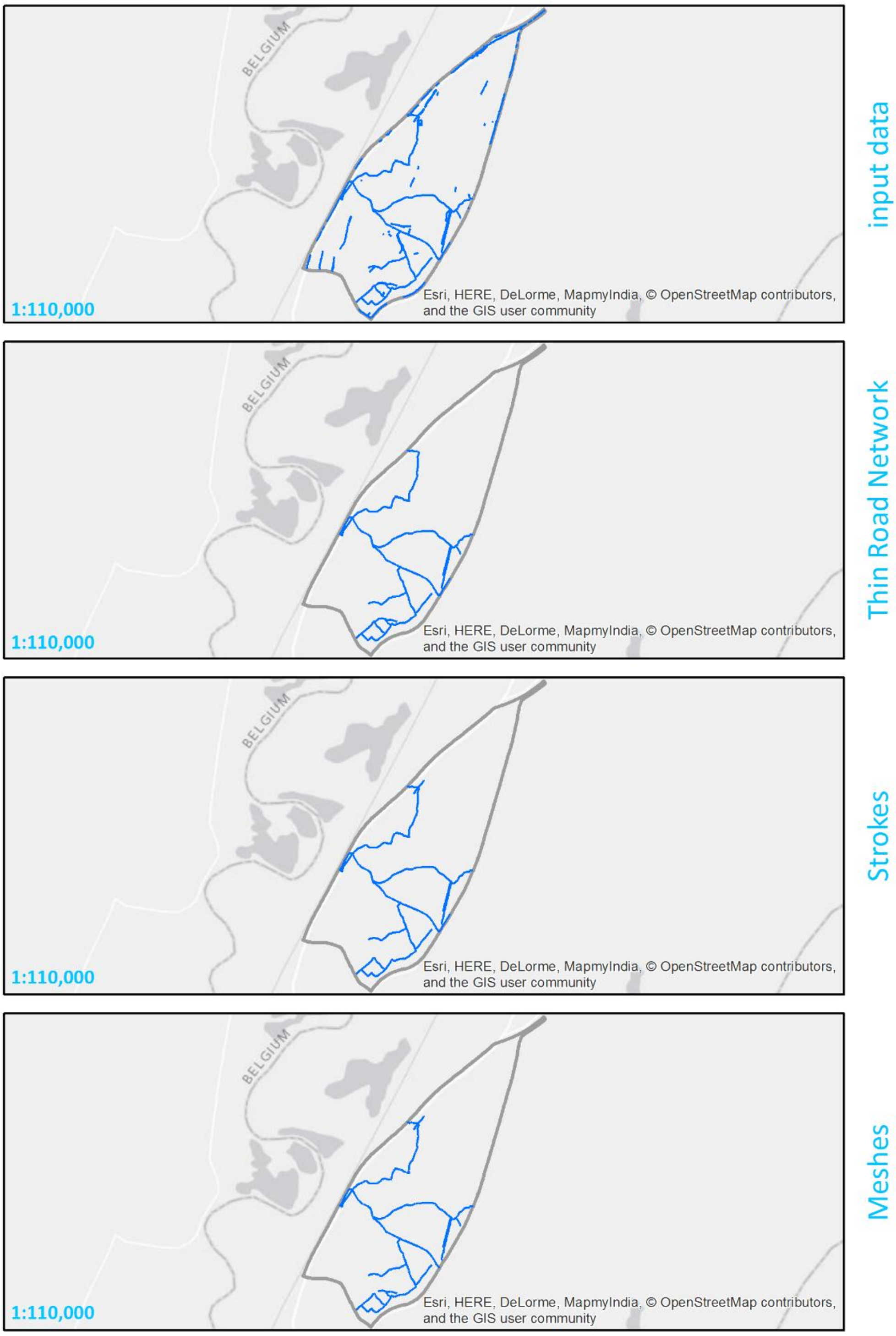


Figure 103: retained largest network after thinning for partition 17

Retained largest network after thinning for partition 49

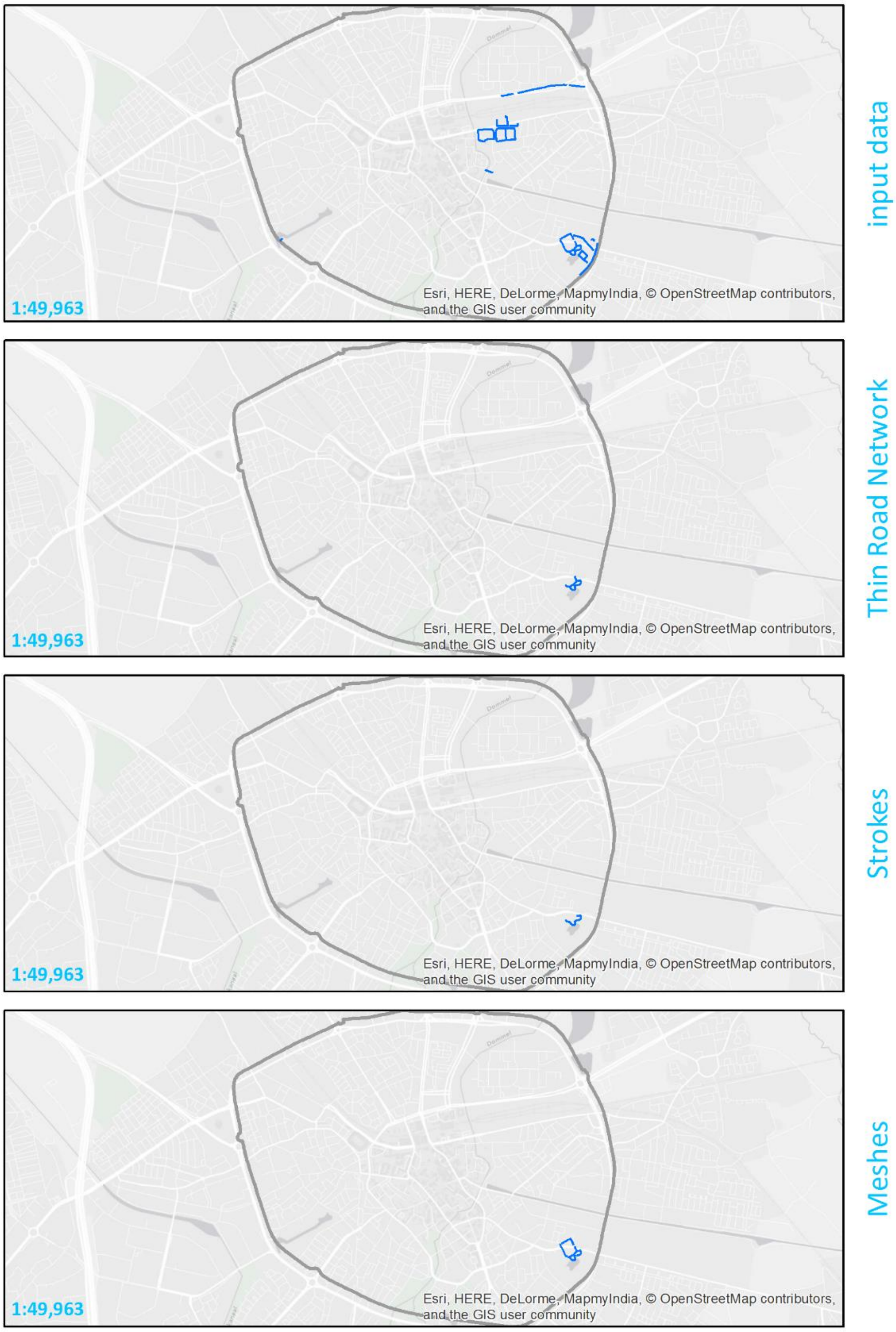


Figure 104: retained largest network after thinning for partition 49

Retained largest network after thinning for partition 93

Figure 105: retained largest network after thinning for partition 93

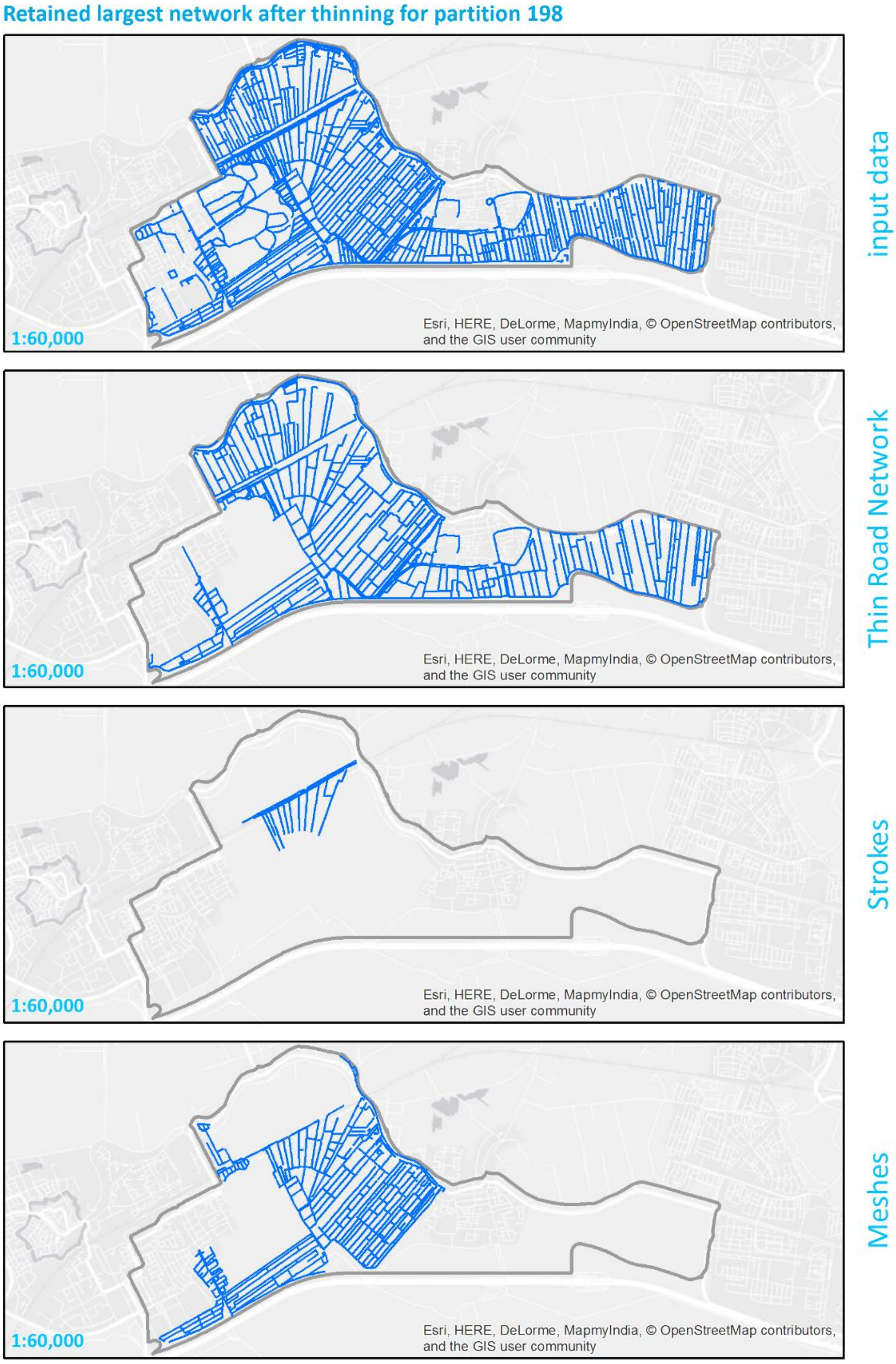


Figure 106: retained largest network after thinning for partition 198

Retained largest network after thinning for partition 217

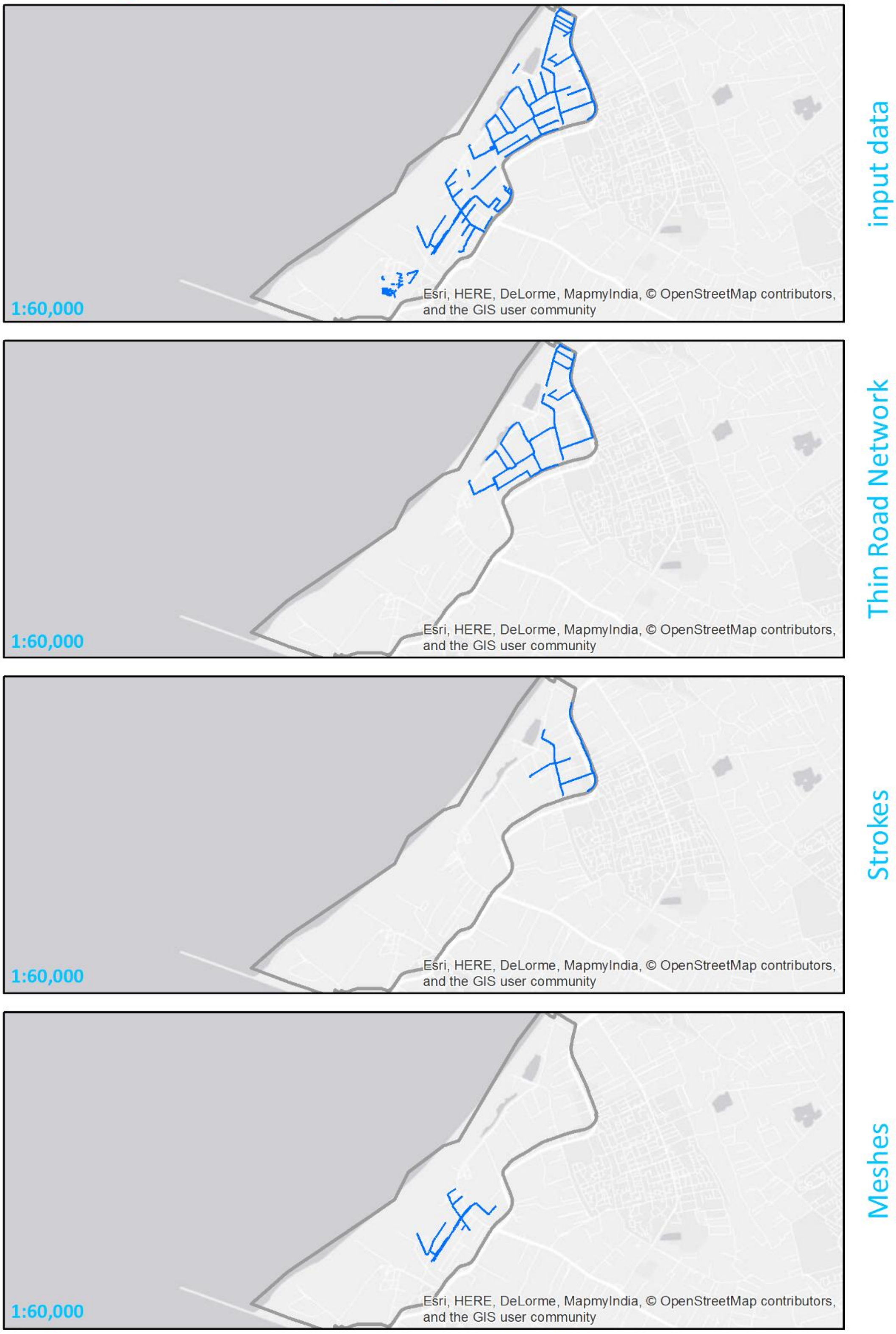


Figure 107: retained largest network after thinning for partition 217

Retained largest network after thinning for partition 260

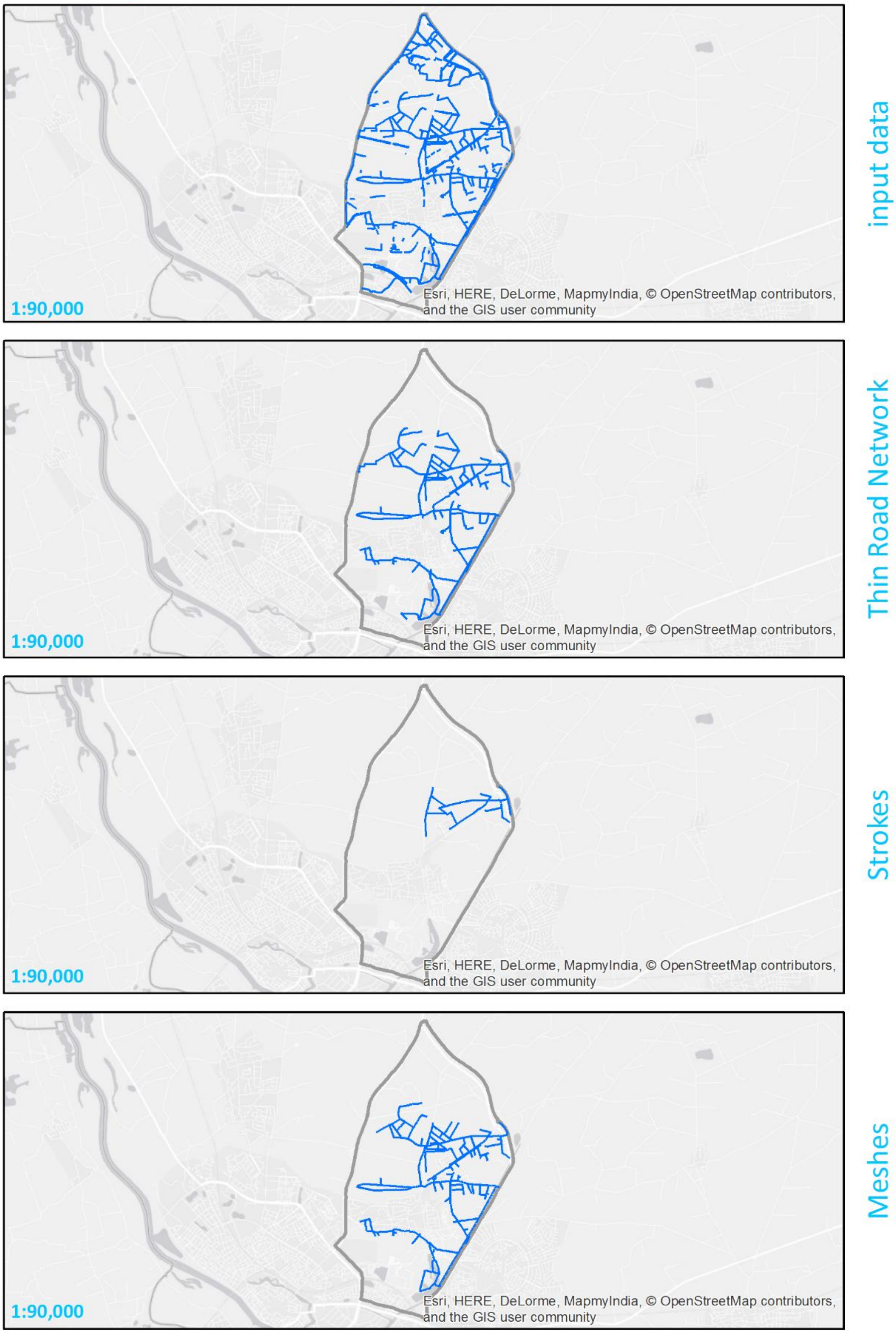


Figure 108: retained largest network after thinning for partition 260

Retained largest network after thinning for partition 298

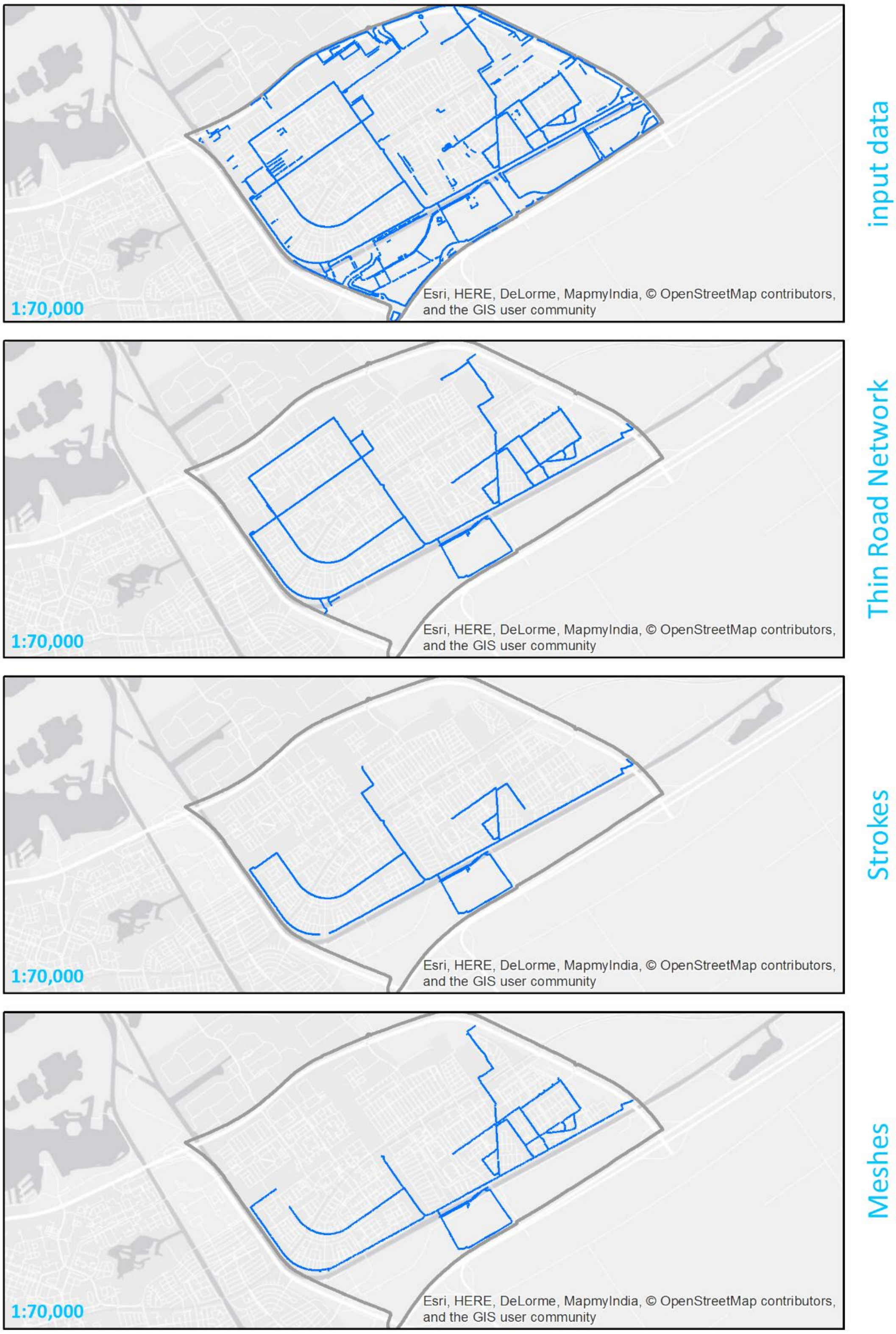


Figure 109: retained largest network after thinning for partition 298

Retained largest network after thinning for partition 345

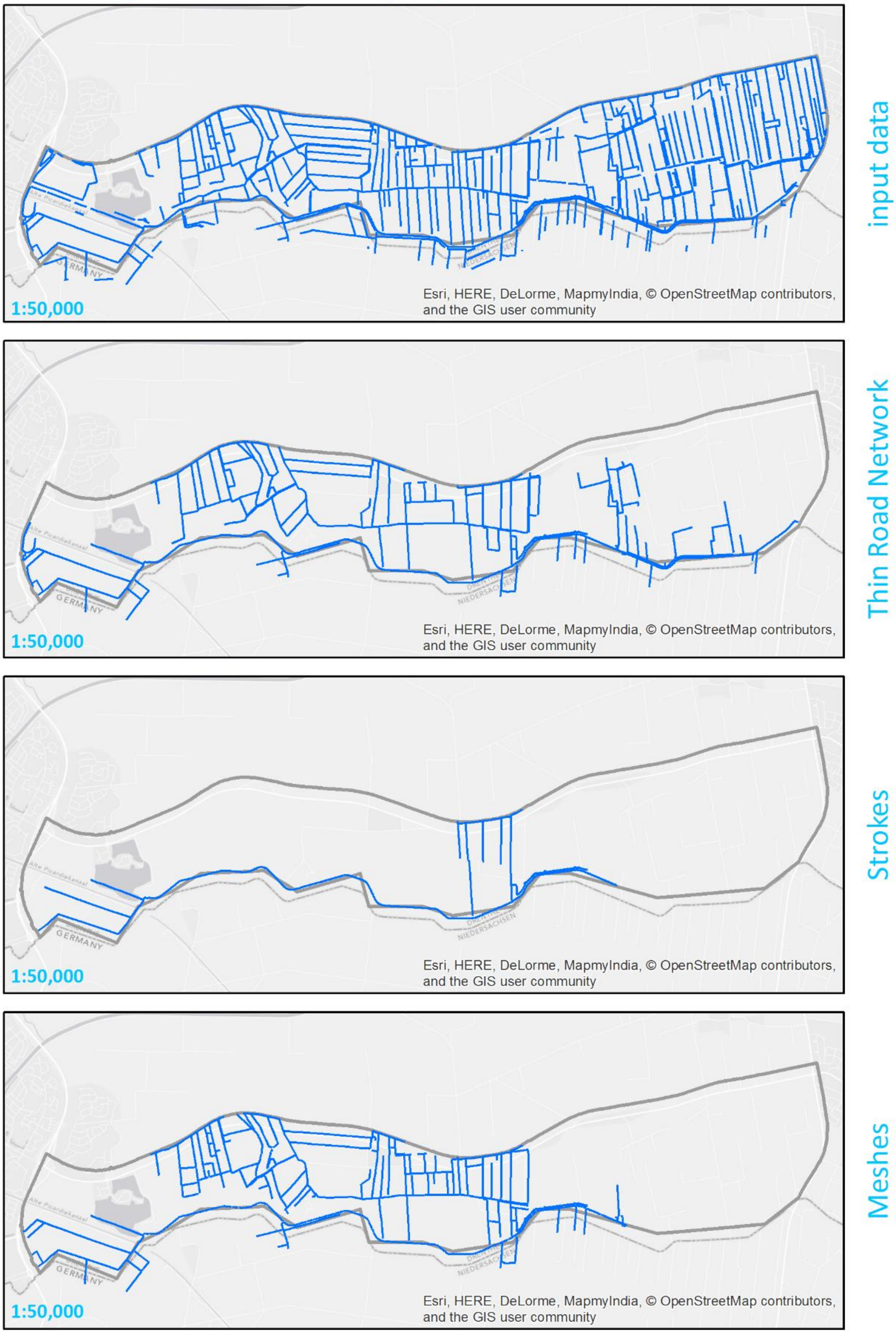


Figure 110: retained largest network after thinning for partition 345

Retained largest network after thinning for partition 403

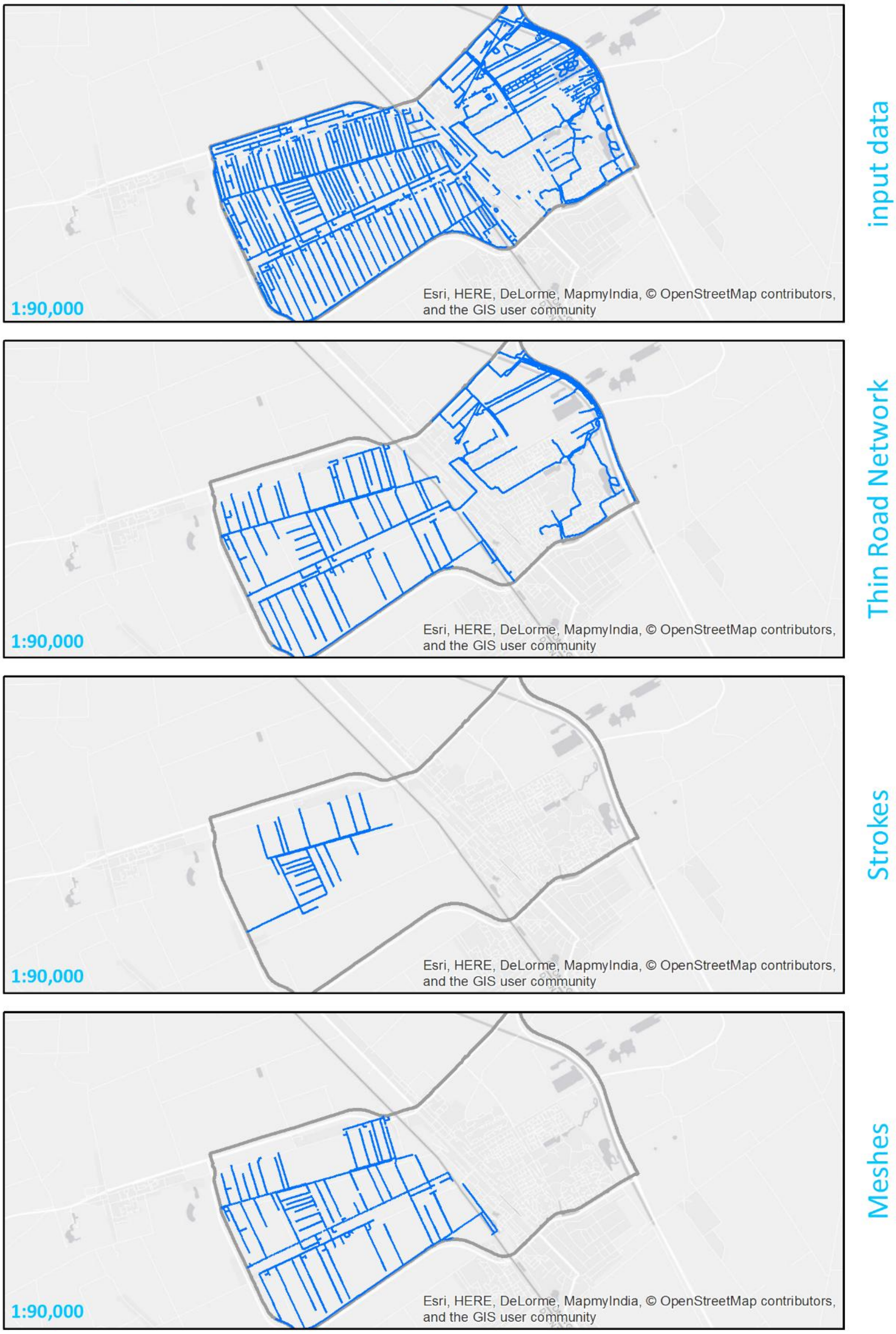


Figure 111: retained largest network after thinning for partition 403

Retained largest network after thinning for partition 442

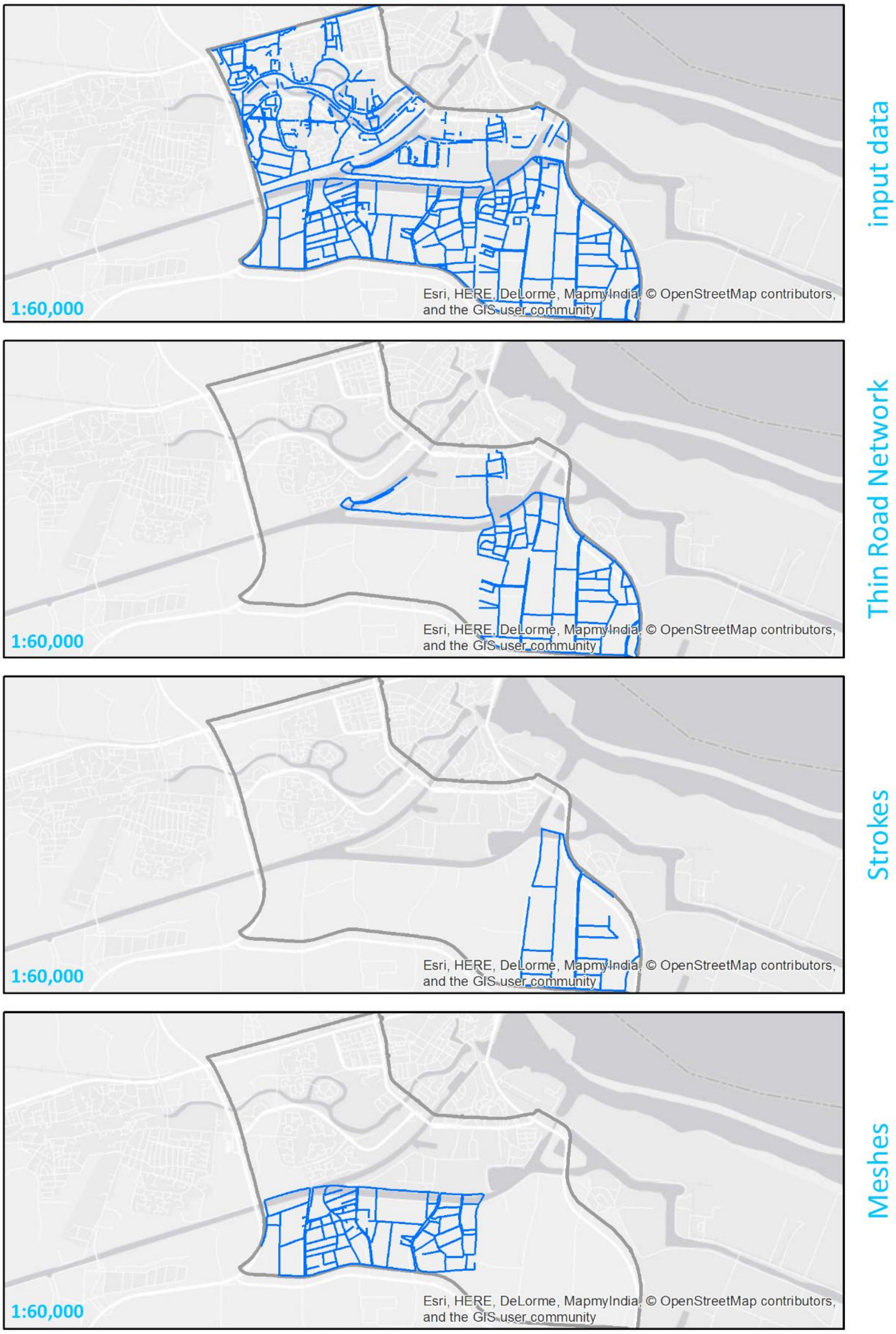


Figure 112: retained largest network after thinning for partition 442