Geographical Information Management and Applications

Master thesis

Quality assessment of the digitized Dutch cadastre of 1832 and its application to study human-land relations of raised bogs

Tom Lonnee

Date:	*1-3-2019*
Student number:	*4111192 (University Utrecht)*
Supervisors:	*Dr. R. van Beek & Dr. M.P.C.P. Paulissen*
Responsible Professor:	*Dr.ir. R.J.A. van Lammeren*

Universiteit Utrecht

TU Delft

University of Twente.

Wageningen UR

Quality assessment of the digitized Dutch cadastre of 1832 and its application to study human-land relations of raised bogs

Master Thesis

Geographical Information Management and Application

Tom Lonnee

4111192

Date: 1st of March 2019

Supervisors: Dr. R. van Beek & Dr. M.P.C.P. Paulissen

Responsible professor: Dr.ir. R.J.A. van Lammeren.

Preface

The thesis that lays in front of you is part of the Geographical Information Management and Application [GIMA] master, commissioned by the University of Utrecht, University of Twente, University of Wageningen, and Delft University of Technology. The topic is a result of earlier experiences with historical geographical data. In earlier courses of the GIMA master I have already applied historical aerial photographs to construct a dataset that stores information of features in history. The use-case was the reconstruction of the landscape of Vlieland in 1981. During this project, I have also started to implement a small quality assessment of the dataset. This quality assessment was the start of the quality assessment part of this final master thesis. Also, my earlier encounter during an internship with *Kadaster International* and a cadastral mapping had an impact on this thesis. I was excited to work with the first version of the cadastre, because some of the principles still apply to the current cadastre.

The use-case of raised bogs was a total new experience, because I have not encountered any related topic during my bachelor's program in Human Geography and Spatial Planning and the GIMA master. The raised bogs use-case was sometimes a struggle, but I want to thank my supervisors, Roy van Beek and Maurice Paulissen, for helping me with advice and constructive feedback. I also want to thank my responsible professor Ron van Lammeren for his guidance during the project. Additionally, I want to thank Johan Feikens and the Fryske Akademy for providing the necessary data for this study. Without this data, this research would not have been possible.

At last, I want to thank my friends and fellow students of the GIMA program for their support, feedback, and remarks on the research. It really helps to discuss the contents of your thesis in an informal setting.

That said, I hope you enjoy reading my thesis.

Tom Lonnee, 24 February 2019.

Summary

The impact of low-quality data on the results is already known among scholars within the field of GIS. However, historians that study history with the help of historical, geographical datasets within a GIS (historical GIS) neglect the quality assessment of their data. The reason for this is related to the complexity of the quality assessment procedure for people that have less knowledge of GIS. The Dutch cadastre of 1832 is a historical, geographical dataset that had become available in a GIS format. However, an assessment of the quality of the cadastre is not available in scientific literature.

The content of the cadastre of 1832 makes it possible how areas were used in 1832. For raised bogs (*hoogveen)* areas, the usage is unknown fefore these areas were reclaimed and exploited. At that time, multiple areas were still raised bogs in Drenthe and Groningen. The availability of the GIS version of the cadastre allows for an extensive analysis with the same settings, which makes it easier to compare the results. This resulted in three aims for this thesis: (1) to assess the quality of the Dutch digital cadastre of 1832; (2) to develop a method for identifying raised bogs in 1832 within a GIS environment; and (3) to study human-land relations for raised bogs with the Dutch digital cadastre of 1832 in Drenthe and Groningen.

The quality assessment of the cadastre is performed with six metrics. The methods to measure these six metrics for the cadastre are different than for datasets that cover the present landscape. Reference data is not available and cannot be used to assess the quality of the cadastre of 1832. Some techniques are inadequate to understand by less experienced GIS users. The quality assessment shows that there is a significant difference (circa 800 square kilometres) between the sizes calculated in 1832 (stored in an attribute) and the size that is calculated within the GIS, where the 1832 attribute is the larger of the two. The attribute accuracy, for historical GIS users the most critical metric, was good. However, this only means that the digitalisation of the dataset was quite well, because the metric only counts the differences between the hard-copy register and the GIS version.

The identification of raised bogs areas could be made with several methodologies. In this study, three are used: a methodology of Spek (2004), a methodology with the BOFEK map (2018) and a combination. All the methodologies used in this research have their inaccuracies. Also, the connection between the cadastre and these raised bogs areas influences the results. Therefore, three options of the select by location tool were used (Contain, Intersect and Within). In this research, the nine studied combinations offer nine different results. Several results do not stroke with existing studies on raised bogs and offer an inaccurate view of the study area.

The last part of the research has focussed on the human-land relations of raised bogs. There is chosen to focus on the differences between five-peat colonial areas and the remaining areas. For the comparison of these areas, the size, land tax value and the land-uses of these areas are further analysed. The results show that three-peat colonial areas (Bourtanger Moor, Hoogeveen and the Smildervenen) have the same land-use distribution as the areas outside the five-peat colonial areas. Also, the size of the parcels and the tax value of these areas were comparable. What stands out was the area of Odoornerveen, which was occupied entirely with the land-use bogs (*veen)*, a low taxation value and large parcels, which show that this area was not reclaimed and exploited yet.

The conclusions of this research lead to recommendations for further research. In this research, the identification methods of raised bog areas are too broad. The methodologies of Spek and BOFEK identifies areas that were raised bogs at some point in history. However, in 1832 some of these areas were already reclaimed. This makes it hard to tell how raised bogs areas were used, before the reclamation. Therefore, the present raised bogs systems could be used because they were certainly raised bogs in 1832.

Table of Contents

List of Figures

List of Tables

1. Introduction

The main themes of this report are the quality assessment of the Dutch digital cadastre of 1832 and the application of the cadastre to study human-land relation of raised bogs areas. The relation between the cadastre and human-land relations, the importance of quality assessment and why raised bogs areas should be studied is highlighted below.

1.1 Problem statement

The application of Geographical Information Systems (GIS) in historical research is a relatively new research field (Dempsey, 2011; Knowles, 2005). With the help of GIS the geography of an area in history can be reconstructed (Dempsey, 2011). The usage of GIS in historical research is called historical GIS (Dempsey, 2011; Knowles, 2005). Historical GIS is not completely new in historical research but has evolved to more advanced tasks. Previously, GIS has been used for visualisation of data or for georeferencing and overlaying historical maps (Gregory & Healey, 2007). Nowadays, more advanced tasks are performed with GIS (Knowles et al., 2008). These tasks and actions could be performed, because of the availability of datasets containing historical data (Gregory, 2008; Gregory & Healey, 2007; Knowles, 2002, 2005).

The datasets that contain historical data are often digitised from older, analogue, data and maps (Gregory & Ell, 2006; Gregory & Healey, 2007). These analogue maps are often far less accurate than the present-day maps. This is caused by technical innovations in the field of surveying, mapping and GIS (Heywood, Cornelius, & Carver, 2011). The present-day mapping techniques (like GPS) are more accurate than the methods that surveyors used in the 1900s. Also, the digitalisation process can cause errors and inaccuracies to the dataset containing historical data (Longley, Goodchild, Maguire, & Rhind, 2005). These errors are created due to misinterpretation of the map by the person that digitises the data. The accuracy and quality of these datasets were already an emerging theme in historical GIS (Knowles, 2005), because of the awareness that low-quality datasets have a huge influence on the final results. Therefore, at least a small quality assessment should be done to analyse the impact on the final results (Girres & Touya, 2010).

Despite the trends observed in 2005, quality assessment is still not common among users of historical GIS (Knowles, 2014). The reason for the lack of quality assessment is related to the experience of the users. The users in historical GIS are often not trained in the usage of GIS and therefore, lack the skills for a proper quality assessment (Knowles, 2014). Another reason that complicates the quality assessment is the lack of alternatives. Therefore, these alternatives could not be used as reference data for the quality control process (Gregory & Healey, 2007; Knowles, 2002). Therefore, the complete research depends on one dataset with historical data. The research cannot be performed without that dataset. A critical reflection that can disqualify the data is therefore often neglected (Knowles, 2014). The same applies to the implications of low-quality data (Agumya & Hunter, 1999; Knowles, 2014).

Historical GIS can be a helpful tool to study relations and patterns in history, such as human-land relations. For example, Knol and Noordman (2003) and Bos (2017) both use a dataset containing historical data, which is analysed with GIS techniques. One of such projects that also study human-land relations, and where historical GIS can be of help, is the Home Turf Project. The Home Turf Project has a specific focus on raised bogs in the Netherlands (Van Beek, Paulissen, & Quik, n.d.). They aim to "design proactive strategies for the sustainable management of bog-related cultural remains" (Van Beek et al., n.d.). Raised bogs were threatened by mainly two processes till World War II (Gerding, 1995). These processes (bog reclamation and peat exploitation) caused that a sizeable part of the raised bogs that have formerly covered the Netherlands has disappeared. But raised bogs are known for high natural value due to the presence of unusual plants and animal species. Therefore, most of the remaining raised bogs are nowadays incorporated in the Natura2000 network of protected areas (European Commission, 2018).

An investigation of the human-land relations at a time when a substantial portion of raised bogs was still present can help to understand how these areas were used and perceived. The Home Turf Project and many other authors have identified the cadastral mapping of 1832 as a helpful dataset for that purpose (Bont, 1997; Grift, 2015; Knol & Noordman, 2003; Van Beek et al., n.d.; Vandenbroucke & Leloup, 2017; Worst, 2012). The mapping of 1832 has recently become available in a GIS format, via HISGIS (Fryske Akademy, n.d.), which offers the opportunity for human-land research over a larger study area compared with the earlier studies (Bos, 2017; Grift, 2015; Knol & Noordman, 2003; Worst, 2012). However, extensive research related to the quality of the GIS variant of the cadastre is not present and cannot be used as a reference. This is surprising when compared to the analogue version of the cadastre dataset. Multiple studies consider the quality of the analogue version (Nijstad, 1982; Veldhorst, 1991). Therefore, this study will also focus on the quality assessment of the cadastre of 1832.

1.2 Research objectives

This research will focus on the application of the Dutch digital cadastre of 1832 for studying human-land relations. Also, the gap in the quality assessment in historical GIS will be addressed. For this research, the following three objectives are formulated:

A. Assess the quality of the Dutch digital cadastre of 1832

The Dutch digital cadastre of 1832 is identified as a helpful dataset for historical studies (van Tussenbroek, 2018). However, information about the quality is not available. To know how the quality of the dataset influences the results of this study, a quality assessment should be performed on the digital version of the 1832 cadastre. The quality assessment will have a specific focus on historical GIS and this creates the potential that parts of the methodology can also be applied on other datasets in historical GIS.

B. Develop a methodology for identifying raised bogs in 1832 in a GIS environment

The methodology will focus on the identification of raised bogs within GIS. There are multiple sources and methods with all their advantages and disadvantages. The goal is to provide some insight into the methods that could be used for the identification of the past geographical location of raised bogs. Also, the impact of these methods on the results will be evaluated.

C. Study human-land relations for raised bogs with the Dutch digital cadastre of 1832 in Drenthe and Groningen

This objective has a primary focus on the attributes of the Dutch digital cadastre and the related patterns for raised bogs areas. These attributes hold the information that is related to the information of a parcel. For example, the name of the owner, the profession of the owner, the land-use, and the land tax value. The application shows the potential of the Dutch digital cadastre of 1832 regarding the study of human-land relations related to raised bogs in Drenthe and Groningen. These two provinces had a large amount of (former) raised bogs in 1832 (Gerding, 1995; Spek, 2004), which are in different stages. Some will be reclaimed and exploited, where others are relatively untouched. The findings will be of value for the Home Turf Project and its research about the raised bogs.

1.3 Research questions

To achieve the objectives, the following research questions are formulated:

1. *What metrics determine the quality of geographical data in Historical GIS and how can these metrics be operationalised to assess the quality of the historical GIS data? (Objective A)*
2. *What is the quality of the GIS version of the Dutch cadastre of 1832? (Objective A)*

3. *How can raised bogs areas in 1832 be identified within a GIS and what is the impact of different methodologies? (Objective B)*
4. *How do small and large raised bogs systems differ in size, taxation, and land-use according to the cadastre of 1832? (Objective C)*

To determine the quality of the cadastral dataset, the metrics that define quality should be investigated. These metrics are already known and well-discussed for present geographical datasets. However, it is unknown if these metrics are suitable for historical, geographical datasets. The second question applies the operationalised metrics on the cadastre of 1832 to assess the quality. The third question relates to the identification of raised bogs in Drenthe and Groningen. A dataset containing the raised bogs areas in 1832 is not available and should be constructed. Multiple sources and methods will be reviewed to see the differences and impact between these methods. At last, the cadastre of 1832 is applied to study the raised bogs areas identified in the third research question. The fourth question focuses specifically on three topics and tries to see the patterns between multiple different areas.

1.4 Relevance

References regarding the quality of the digital version of the Dutch cadastre of 1832 do not exist. But due to the impact of low-quality data on the results, an assessment of the quality could help to confirm these results. This research aims to offer an assessment of the quality of the cadastre of 1832. The quality should be considered, when the cadastre of 1832 is used in other studies. Also, existing research shows that quality assessment is not performed by users of historical GIS (Gregory & Healey, 2007; Knowles, 2014). Therefore, this research also aims to develop a methodology for the quality assessment that can (partly) be applied to other historical, geographical datasets.

More specifically, the Home Turf Project can use the results of the quality assessment, when they use the cadastre of 1832 in their research. The same goes for the application of the methodology of the quality assessment on other historical, geographical sources. The application to the raised bogs is also relevant for the Home Turf Project. The Home Turf Project studies raised bogs, which are often not visible anymore in the current landscape. The different methodologies for the identification of raised bogs can help them to identify where these areas are found in Drenthe and Groningen.

The application of the cadastre on the raised bogs areas provides information about these areas for the Home Turf Project. In this research a large area is studied at once. This makes it easier to compare the findings, because the same method is applied. The study of a larger area is a current gap in research, because most studies about human-land relations focus on a relatively small study area. Examples of these small study areas are a valley in Gelderland (Bont, 1997; Grift, 2015), a municipality of Heinoo (Knol & Noordman, 2003), four towns in Friesland (Worst, 2012), the Weerribben (Reeskamp, 2013) and Twente (Van Beek, Maas, & Van Den Berg, 2015) and Hennaarderadeel (Bos, 2017).

1.5 Reading guide

The structure of the report is as follows. The first chapter contains an introduction on the problem and offers the research questions stand central in this research. The next chapter contains theory and literature about quality and raised bogs. The third chapter holds the methodology of this research. The methodology chapter will also have a description of the study area and the cadastral mapping of 1832. The next two chapters will focus on the results of the study. First, the results of the quality assessment, which are followed by the results of the analysis of raised bog areas. The results are discussed in the following chapter, as well the limitations of the research. The research is concluded with a chapter containing the answers on the research question and the recommendations of further research.

2. Theoretical Framework

The theoretical framework will start with an explanation of quality assessment. What is quality assessment and what defines the quality of a geographical dataset? The quality assessment is also linked to historical GIS. The second part of the theoretical framework will focus on the development and disappearance of raised bogs, as they are a critical element in this study.

2.1 Quality Assessment

As mentioned earlier, the quality assessment in Historical GIS is not a common aspect. Users assume that the quality of the used datasets is good. Implications of bad data are neglected, but bad data can greatly influence the outcomes of the research (Knowles, 2014). Also, a lower data quality hinders the usage of the data in other studies (Van Oort, 2006). Many studies have studied quality of geographical data. Most of the distinguished metrics are also implemented in a standard. The standard geographical information – data quality (ISO19157:2013) is one of the most widely used standards (Van den Berg, 2018). This standard is created by the International Standardisation Organisation [ISO] (International Organization for Standardization, 2013). The ISO described in the standard the guidelines for evaluating the quality of geographical data. These guidelines include the components of geographical data and how these can be measured.

The ISO 19157 standard distinguishes two main categories relating to the quality of geographical information. The two categories are the internal quality and the external quality (International Organization for Standardization, 2013). The latter is also known as pragmatic quality in other studies (Bordogna, Carrara, Criscuolo, Pepe, & Rampini, 2014; Van den Berg, 2018). But both the external and the pragmatic quality are related to the application of the data in a case study. The difference between the external and pragmatic is related to the experience of the user. The experience is included in the ISO standard, but is not in the pragmatic quality definition of Bordogna et al. (2014). This is due to the fact that Bordogna et al. (2014) focus on Volunteered Geographical Information. By Volunteered Geographical Information the user experience is more important than by more traditional datasets (Bordogna, Carrara, Criscuolo, Pepe, & Rampini, 2016). However, the structure of the report will follow the main categories of the ISO 19157 standard (2013) and Van Oort (2006).

2.1.1 Internal quality metrics

The internal quality is a concept that *"document inherent characteristics of geographical data that will be useful for every user to evaluate their ability to fulfil their application requirements"* (Bucher, Falquet, Metral, & Lemmens, 2016, p. 133). Bucher et al. (2016) argue that data producers always should distribute the data together with the description of the internal quality. This help the users to assess the usability of the data for their application (see 2.1.2). The metrics that are part of the internal quality differ per source. Van Oort (2006) illustrates this in his PhD research on geographical information quality at Wageningen University. Van Oort compared multiple sources relating to spatial data quality and classified the metrics in three categories. Van Oort observed that eleven elements were mentioned at least once in the five studies (Table 1).

The comparison of Van Oort (2006) does not include the ISO 19157 standard, because the ISO standard was still in development at the time. However, the contents were already known, because Van Oort refers to the standard in his PhD research. Five of the eleven metrics of his list are not included in the ISO standard. Three metrics (variation in quality, meta-quality and resolution) are encountered as sub-elements rather than individual metrics (Van den Berg, 2018). Semantic quality is also not included in the ISO standard. However, the semantic quality is only mentioned by the ICA, short for International Cartographic Association, and the CEN, Comité Européen de Normalisation. However, Van Oort (2006) mention in his research that the CEN will adopt the ISO 19157 standard, when the development is completed. Therefore, it is likely that the semantic accuracy will disappear as a metric of quality. This makes lineage the only metric that is not included in the ISO 19157 standard.

	Aronoff (1989)	USA-SDTS (1992)	ICA (1995)	CEN TC287 (1998)	ISO TC211 (2002)
Lineage	S	S	S	S	S
Positional accuracy	S	S	S	S	S
Attribute accuracy	S	S	S	I	S
Logical consistency	S	S	S	S	S
Completeness	S	S	S	S	S
Semantic accuracy	-	-	*S*	*S*	-
Usage, purpose, constraints	S	M	-	S	M
Temporal quality	S	M	S	S	S
Variation in quality	-	I	I	S	I
Meta-quality	-	I	I	S	I
Resolution	S	I	I	I	M
S =	explicitly recognised in the spatial data quality section of the metadata				
M =	explicitly recognised as an element				
I =	in another section of the metadata and implicitly recognised as an element				
- =	not recognised				

Table 1 - Internal quality metrics of geographical data (Source: Van den Berg, 2018; Based on Van Oort, 2006).

Based on Van Oort (2006) and the ISO (2013) seven metrics can be distinguished for the internal quality. This included the original six ISO metrics and the lineage metric. Below, a description of the metrics is described based on the research of Van Oort (2006, pp. 14–18).

1. Lineage

Lineage can be described as the history of a geographical dataset. It supplies a description of the source material, and the data obtained from the source material. It also includes the further processing steps.

2. Positional accuracy

The positional accuracy is the degree of correspondence between the coordinates of the data and the real world. The X and Y coordinates stand for the horizontal positional accuracy. The Z coordinate represents the vertical positional accuracy. However, the Z value is sometimes treated as attribute accuracy, because of an attribute that holds the height or depth value. Another distinction is between relative and absolute positional accuracy. The latter compares the coordinates of the features of one dataset with a reference dataset. The relative positional accuracy is related to the other data in the same dataset. An example to illustrate this:

Building A has coordinates 123 and building B has coordinates 987 in the real world. The distance between the two buildings is exactly 1 kilometre. In a dataset, building A has coordinates 456, and building B has coordinates 321. But the distance between the two is still 1 kilometre. This means that the absolute positional accuracy is low, because the buildings are on a different location. But the relative accuracy is sufficient, because the space between them is still the same as in the real world.

3. Temporal quality

This metrics is called different in all the five sources that are used by Van Oort (2006). But all contain information that is related to time. Not all the sub-elements measure errors, and this is the reason that Van Oort does not use the word accuracy, but temporal quality. Some elements that are distinguished are information about the last update, the correctness of the order of events and the so-called temporal lapse. This last element is the average time between a change in the real world and its representation in the data.

4. Attribute accuracy

The attribute accuracy is the accuracy of all attributes that are not related to the positional and temporal attributes. These attributes could be measured in different measurements scales (nominal, ordinal, interval, and ratio). In every scale, a different method could be used to measure accuracy. The nominal scale could use an error matrix, and for ratio variables, it could be measured with the root mean square error (RMSE) (Gabriela, Stefan, & Aurelian, 2010; Van Oort, 2006).

5. Logical consistency

The logical consistency focusses on the relations between the data of features, in case of the cadastre the features are the parcels. The most principal element of logical consistency is the topological relation. The topological relations provide insight if all the features are aligned. When gaps and overlaps exist then the relations between the features are not well defined. These gaps and overlaps hinder the analysis and provide more inaccurate results.

6. Completeness

The completeness of a dataset is related to the degree of absence of data, that should be present in the dataset, or the presence of data that should not be present in the dataset. Van Oort (2006, p. 16) provides a clear example of completeness relating to hospitals. The detection of errors could only be measured when there is known what does and what does not belong to the dataset. Also, the definitions that are used should be known. These definitions should be documented for the usability of the dataset. Unknown values in a GIS also influence the completeness of a dataset.

7. Usability

Usability is a metric that is not consistently used as an internal quality metric. Where Van Oort (2006) and the ISO (2013) count usability to the internal quality, Bucher et al. (2016) argue that the usability is also an external component (see paragraph 2.1.2). The discussion is caused by the fact that some elements apply to both perspectives. However, authors agree that metadata and documentation are essential for assessing the usability (Bucher et al., 2016; Van den Berg, 2018; Van Oort, 2006).

For the internal quality, the description of Van Oort will be followed. The usability, or as Van Oort called it: ‘usage, purpose, and constraints’ focus on the details of the content. An example of this is supplied above; documentation that include the definitions of attributes. This allows the user to use the attributes more easily. Good documentation leads to less time and costs spend on understanding the dataset. These resources could be spent on other aspects of the research, which increases the usability.

Van Oort also mentions the costs and constraints as an internal quality aspect. However, these elements are more in line with the description of Bucher et al. (2016) and Bordogna et al. (2016). These elements are therefore added to the external quality.

2.1.2 External quality metrics

The ISO classifies external quality as the fitness-for-use of the data (International Organization for Standardization, 2013). This means that the external quality is closely related to the use-case. The most important aspect according to Bucher et al. (2016) is the ability to discover and re-use the data in the research or project. The discovering is related to finding the data. For example, is the data findable for the public or only by a specific department or company? The re-use of data is related to the financial and legal constraints. Sometimes a fee needs to be paid, or can the data not be used within an application that will be sold to companies (Van Loenen, 2006). Another aspect is the practical ability to combine the data with other data. This is often related to format of the dataset (Geiger & Von Lucke, 2012). Some formats (like PDF files) can be opened by multiple software, but shapefiles can only be processed by geo-software (Löwenberg, 2014). These constraints lower the external quality of a dataset.

The last aspect is even more related to the use-case and the relating processing time (Bucher et al., 2016; Turner, 2002; Van Oort, 2006). How much time is needed to make the dataset suitable for your analysis and use-case? One element is related to the time needed to solve errors (Bucher et al., 2016; Van den Berg, 2018). These errors are found during the analysis of the internal quality and should be solved by the users, when the quality is not sufficient. Another element is for example reclassification of attributes and the needed level of detail. Some datasets have collected some attributes on a detailed level. However, for some users this level of detail is not necessary or not practical. They must reclass the data and the time needed for this influences the external quality.

	Metric
Internal	Lineage
	Positional accuracy
	Temporal accuracy
	Attribute accuracy
	Completeness
	Usability
External	Discovery and re-use of data
	Data format
	Use-case

Table 2 - Quality assessment metrics (Bucher et al., 2016; International Organization for Standardization, 2013; Van Oort, 2006)

2.1.3 Quality assessment of the digitized datasets

As explained in the problem statement (paragraph 1.1), the field of Historical GIS depends highly on digitised data. The sources in Historical GIS are often old maps that are not available as a format that can be directly used in a GIS (Gregory & Healey, 2007; Knowles, 2002). These analogue versions are made on paper and are saved in an archive. To enable these paper maps for usage in a GIS, the data should be digitised (Gregory & Healey, 2007).

This digitalisation can be done in multiple degrees (Longley et al., 2005). All these methods start with a high-resolution scan of the analogue data (Olah, 2009). The high-resolution scans are geo-referenced, which means that coordinates are added to the scans (Olah, 2009). Objects on the scan have now the same coordinates as their real-world counterparts (Longley et al., 2005). After the georeferencing procedure, the interpretation can start. There are two methods for this interpretation, the manual/visual and an automatic procedure. The visual interpretation means that a human makes an interpretation of the contents of a map. For example, Spek (2004) looks at old military maps and tries to find areas that have a name that ends with ‘veen’ or ‘moer’. The other method is to let software analyses the data. Therefore, attribute data is needed, which georeferenced maps do not have. Therefore, an extra processing step should be performed.

The data should be processed to shapefile that only represents the features that you want to analyse. This can be done by drawing the features of the map to the shapefile. The shapefile acts as the layer that lays on top of the georeferenced image. The features that are drawn are stored in the shapefile. To these features the attributes could be linked (Olah, 2009). In case of the cadastre, this means that the parcels are drawn from the georeferenced map to a geodatabase layer. The next step was the digitalisation of the register that stored the attributes of the parcels, however more about this process can be found in paragraph 3.2.

Literature about the quality assessment of digitised GIS data is not widely available. However, the digitalisation of older, analogue source material has its consequences for the quality assessment. The features that are represented are often no longer recognizable in the real world. They have vanished or are altered over the years (Bont, 2004). Also, the digital version of the analogue version is unique. Alternatives are not available that contains the same information of the same areas (Gregory & Healey, 2007; Knowles, 2002, 2014). This means that a common method of quality assessment, a comparison with reference dataset, cannot be performed (Van Oort, 2006).

This means that other methods should be used to assess the quality of digitized datasets. One method is using the assumption that the original source is correct and complete (Knol & Noordman, 2003). This provides a reference for comparison with the digitised output. This provides information if the digitalisation process is correctly done. The analogue map should be compared visually with the digitised version. However, this is a very time-consuming process, even when a sample is used.

Another method is to just focus on the quality metrics that are related with the digitised dataset (Agumya & Hunter, 1999; Pipino, Lee, & Wang, 2002; Wang & Yin, 1997). This method assumes that all the features that are mapped are correct and looks at missing attributes, topological relations, and the usability of the dataset for your project. The downside from this method is that only a few aspects of the quality are reviewed. For example, the positional accuracy is completely neglected in this method.

2.2 Raised bogs

This paragraph will focus on raised bogs, which should be analysed with help of the cadastre of 1832. Goal of this paragraph is to highlight the development of bogs and why the raised bogs have disappeared, because not all readers will be familiar with this topic.

2.2.1 The development of raised bogs

Bogs (*hoogveen*) and fens (*laagveen*) are both part of the peat soils (*veengronden*). These peat soils consist of a layer of dead vegetation (Casparie, 1993; Pleijter, 2004). In the soil science, three types of peat soils are recognized. These three types of soil are called: eutrophic fens, intermediate fens (*overgangsvenen)* and oligotrophic bogs (Bont, 2009; Grift, 2015; Zagwijn, 1986). The classification is based on the nutrient content of these bogs. Eutrophic fens contain the most nutrient, and the oligotrophic bogs contain the lowest amount of nutrient. The intermediate fens is the in-between variant (Berendsen, 2008). These bogs types follow a vertical sequence with the eutrophic fens at the bottom (Zagwijn, 1986). The intermediate fens grow on top of the eutrophic, where the oligotrophic follows the intermediate stage (see figure 1).

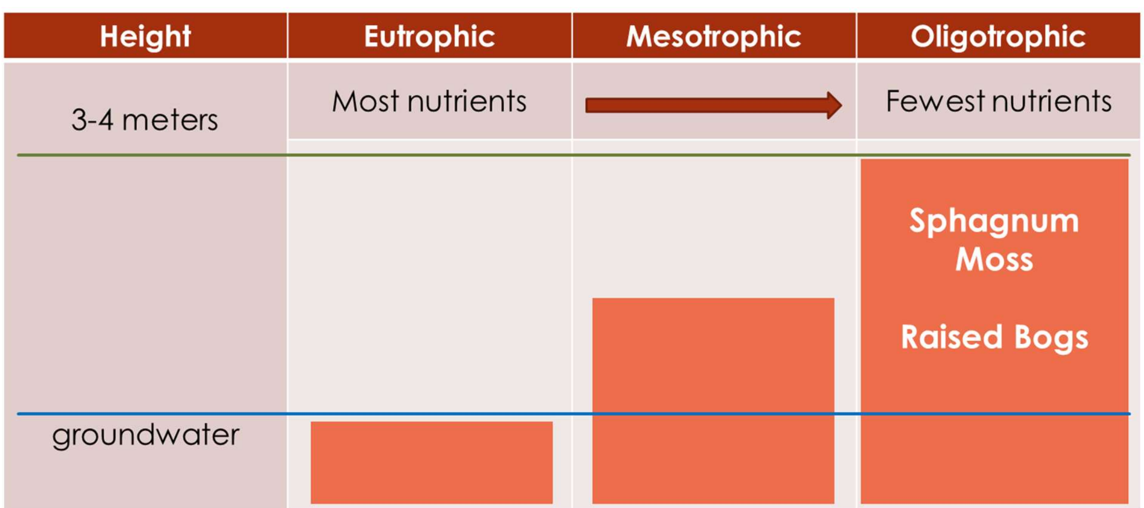


Figure 1 - The sequence of bog development. The development of bogs through time (from left to right) (based on Bont, 2009).

The development of bogs starts at relatively low and wet areas, in comparison to the surrounding areas. When these areas lay below the groundwater level, the creation of eutrophic bogs starts. When the vegetation dies, it creates a layer of soil. On top of this layer other vegetation starts to grow. After some time, this vegetation dies as well, and the process is repeated. The soil that is created below the groundwater level is called fens (Zagwijn, 1986). Eventually, the layer of dead vegetation is so thick, that the bogs are above the groundwater level (Bont, 2009; Zagwijn, 1986). The higher the soil is above the groundwater level, the fewer nutrients the soil contains. This is due to the fact that rainwater contains far fewer minerals than groundwater (Kekem, Hoogland, & Horst, 2005). The vegetation adjusts itself to the conditions.

In eutrophic fens, vegetation occurs that can grow in a nutrient-rich environment. The vegetation feeds itself with both water from the ground as from the rain. When the soil lays above the groundwater level, more vegetation with large roots will follow (Grift, 2015). The large roots enable the opportunity to still reach the groundwater for their nutrients. When the bogs are more than a few metres thick, vegetation cannot reach the groundwater anymore. The vegetation is replaced with sphagnum moss. This moss can retain the water from the rain very long period (Gerding, 1995; Kekem et al., 2005). The oligotrophic bogs with sphagnum moss are the areas that are nowadays called raised bogs.

These raised bogs do not only evolve from the eutrophic fens. They could also be created on places, mostly sand soils, that lay relatively low compared with neighbouring areas. These sand soils lay often high above the groundwater, which resulted in the grow of sphagnum moss in these depressions (Jong, 2009). This resulted in raised bogs areas, that could exist without a connection with the eutrophic fens.

Originally, it was thought that raised bogs only occur in the higher parts of the Netherlands (Bont, 2009; Schouwenaars, Esselink, Lamers, & Molen, 2002). Raised bogs also occur in the lower, western parts of the Netherlands. These bogs are already millennia old and at that time the sea level was several metres lower than nowadays. At that time, the raised bogs are developed, the same way as described above. When the sea level rises in the Holocene, these raised bogs were 'drowned' (Schouwenaars et al., 2002). For a reconstruction of these 'drowned' raised bogs, see Zagwijn (1986). However, the appearance of raised bogs areas in the lower parts of the Netherlands is not only caused by the rise of the sea level. The raised bogs areas in the western part of the Netherlands, were the first to be exploited and reclaimed. The peat, which was several metres thick, have completely vanished and this now results in a soil that lays several metres lower than surrounding non bogs soils (Homburg, 1991).

But also, the opposite is true. Fens due not only occur in areas below sea level. There are also fens in the higher parts of the Netherlands. The grow of sphagnum moss have not continued in these areas (Zagwijn, 1986). Therefore, a distinction of raised bogs and fens based on the Dutch sea level (NAP) is not possible.

2.2.2 The disappearance of raised bogs

The previous paragraph focusses on the development of raised bogs. However, the raised bogs that formerly covered a significant part of the Netherlands have disappeared (F. De Vries, Hendriks, Kemmers, & Wolleswinkel, 2008). Authors do not agree on the exact amount of raised bogs that have disappeared. However, the general trend is that more than 95 percent of the raised bogs that formerly covered the Netherlands have disappeared (Bont, 2009; Gerding, 1995; Schouwenaars et al., 2002). The disappearance is caused by two processes: bog reclamation and peat exploitation.

Bog reclamation is done since the tenth century. The first reclamations of the fens and raised bogs took place in the provinces of North- and South Holland. The reclamation took primary place for agricultural purposes (Van Beek et al., 2015). More agricultural land was needed in order to feed the growing population of the Netherlands (van Zanden & van Leeuwen, 2012; Wolff, 1992). The reclamation shifted roughly from the coastal areas towards in-land areas and from the south (Flanders) to the north (Drenthe) (Van Beek et al., 2015, p. 11).

The growth of the population was the start of large-scale peat exploitation. More people start demanding fuel for their daily life (Gerding, 1995; Wolff, 1992). The disappearance of forest, because wood was the primary energy source, leads to a shift to the usage of peat. Peat is the soil that can be found in raised bogs and was just as efficient for energy production. The demand for peat leads to the foundation of various peat colonies in the Northern part of the Netherlands (Gerding, 1995). The population of these colonies worked primarily for the peat exploitation (Gerding, 2009). When an area was fully exploited, the areas were mostly modified to make it suitable for agricultural processes (Van Beek et al., 2015).

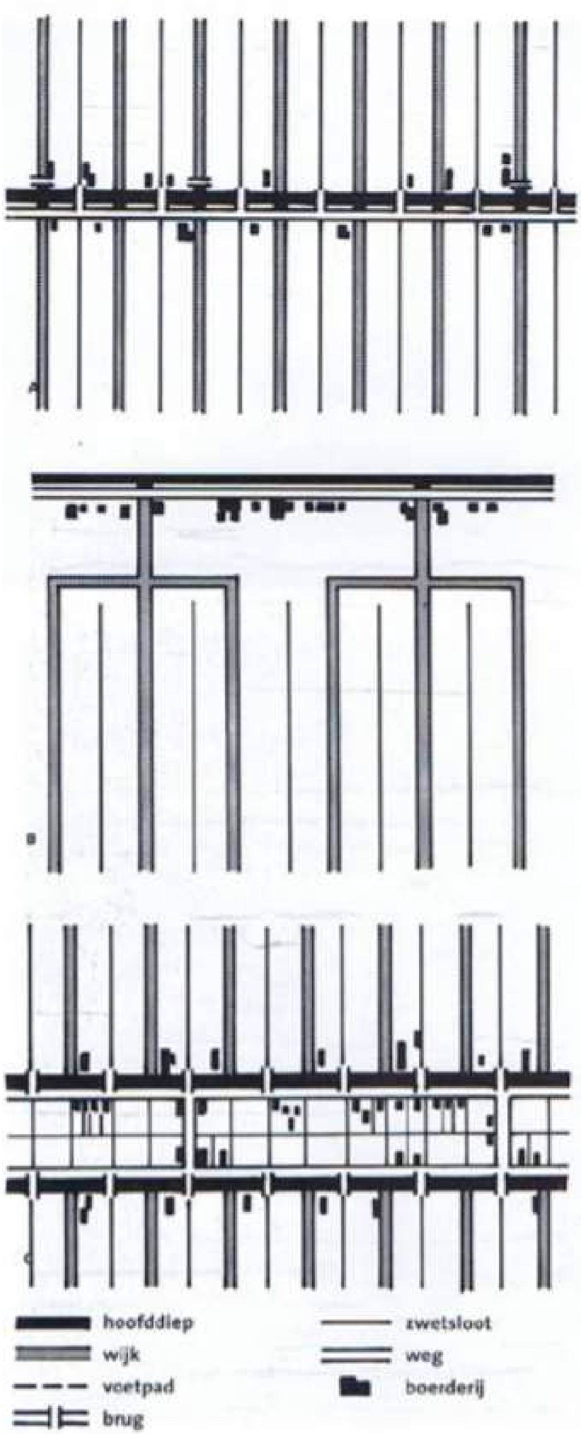


Figure 2 - Three types of settlement structures in the peat colonies according to Stol (1997). Above with one canal, in the middle the trident structure and below the double canal structure

The exploitation and reclamation had an enormous impact on the environment. The reclamations started mostly from a central axis. This was mostly a canal, natural valley or, in some occasions, a road (Bont, 2009; Gerding, 1995; Van Beek et al., 2015). At right angles of this axis, narrow ditches were dug. These ditches divide the land in small, but long plots. The ditches and the canal were necessary elements in the reclamation and exploitation process. The canals serve both as drainage, but also as a transportation method (Gerding, 1995). The farmstead was built at the front of the plot, and from there the area was further reclaimed (Grift, 2015; Van Beek et al., 2015). This leads to a linear settlement pattern (Figure 2). The long plots of lands are still visible in the parcellation pattern of some areas in the Netherlands (Stol, 1997). However, some areas are also influenced by land consolidation, because the ditches were no longer necessary, and some plots of lands could be reallocated together. This results in larger parcels and also to more financial profits for the farmers (Gerding, 1995).

Nowadays, only a small part of bogs has survived the two main processes. The areas that remained are almost all protected nature reserves (Groenewoudt, 2009, p. 162). The Natura2000 policy protect these areas due to the high natural value. The high value is related to the presence of unusual plant and animal species, like the sphagnum moss (European Commission, 2018). The raised bog landscape also houses cultural remains, that can essential for assessing the human impact on these areas.

3. Methodology

3.1 Study area

The provinces of Drenthe and Groningen in the north of the Netherlands are selected as the study area of this research. The study area is larger than in similar studies that involve human-land relations (Bont, 1997; Grift, 2015; Knol & Noordman, 2003; Worst, 2012). The study area is used in all elements of this research. The quality assessment will be performed on datasets that cover these two provinces. Also, the methodological steps for raised bogs distinction and the study of the human-land relations will focus on this area. Cultural and geographical arguments support the choice for Drenthe and Groningen.

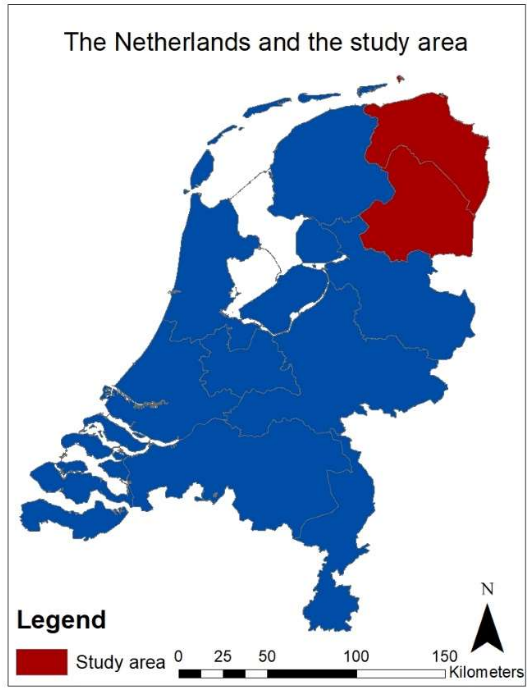


Figure 3 - The location of study area in the Netherlands

Drenthe and Groningen have both a relatively large amount of (former) bogs compared to other provinces. Only the province of Friesland has similar amounts of bogs (Gerding, 1995). The large amount in Drenthe and Groningen can be partly explained with the presence of the Bourtanger Moor. The Bourtanger Moor was the most extensive system of bogs in north-west Europe (Casparie, 1993). Other areas, small and large, were located in these provinces as well (G. De Vries, n.d.; Jong, 2009; Pleijter, 2004). The high amount of (former) raised bogs makes Drenthe and Groningen a suitable location for this research.

Another argument is related to the importance of peat exploitation and bog reclamation in Drenthe and Groningen. The exploitation and reclamation started already in the 16th century. Many companies were involved in the exploitation of peat for energy production. This was a profitable business that lasted until the second World War (Bos, 2017; Gerding, 1995, 2009). In 1832, both reclamation and exploitation were present in various stages in Drenthe and Groningen. The situation in 1832 shows a diverse area regarding reclamation, peat exploitation history, and land use patterns. Some bog areas have already been fully exploited or reclaimed, other areas only partly so, yet other bog areas are still relatively untouched. In comparison, Friesland is almost fully exploited in 1832 (Gerding, 1995) and does not have this diversity. The diversity in stages of exploitation and reclamation offers an extra opportunity to analyse patterns regarding human-land relations.

3.2 The cadastre of 1832

The Dutch cadastre of 1832 is the main source in this study. The quality is assessed, and the cadastre is used for the analysis of raised bogs. This chapter will describe the origin of the Dutch cadastre and how the cadastre is digitalised. Main source for the origin of the cadastre will be Veldhorst (1991). Veldhorst focusses on how the cadastre can be used in other study areas, like historical soil studies and history of farming techniques.

The Dutch cadastre of 1832 is the first nation-wide cadastral mapping of the Netherlands (Veldhorst, 1991). The final measurements and registry were performed in 1832, but it already started in 1810. In this year, the area that we now know as the Netherlands were part of the French Empire of Napoleon (van der Burg, 2007). The French started a nation-wide system for the collection of taxes, the cadastre. This system was adopted by our king when the French left the Netherlands. In 1832, the cadastre was a complete registration of soil ownership and the corresponding taxes that should be paid. The cadastre consisted of both maps and registers.

A surveyor made the maps. The first thing a surveyor did was to divide the municipality into sections. These different sections were mapped on a separate map, which was called the overview sheet (*verzamelblad*). For all these sections, more detailed maps were created (*Minuutplans*). These detailed maps showed the boundaries of the parcel and the corresponding number of that parcel. The municipality name, the section and the parcel number provided a unique combination, that is used in the registration books. The main registry document was the '*Oorspronkelijke Aanwijzende Tafel der grondeigenaren en der ongebouwde en gebouwde vaste eigendommen* (…)', this document will be mentioned shortly as OAT. The OAT was a registry with all the parcels of the cadastral municipality. For every parcel, information about the owner and the parcel is noted. An example of parcel information is the land-use and the size of the parcel. Of course, also the value and the taxes that should be paid are noted (Veldhorst, 1991).

Veldhorst ends his article with a critical reflection on the contents of the cadastre, especially the valuation and taxation part. The cadastral surveys were conducted per cadastral region. The results and the valuation of the land are therefore specific for these regions. The valuation of the lands differs between the various regions (Veldhorst, 1991, p. 26). The same lands were valuated differently in different regions. This also influences the taxation of the parcels. The argumentation of the different valuation is covered in so-called base documents. However, these base documents are mostly destroyed. Therefore, the source that should ensure uniformity has vanished. Veldhorst (Veldhorst, 1991, p. 25) argues that the valuation of the parcels should be used for larger regions without a critical reflection of the user. This observation can have severe consequences for the final analyses.

These maps and OATs are only available as hardcopy maps. However, the HISGIS project, hosted by the Fryske Akademy, digitalises the Dutch cadastre of 1832 (Fryske Akademy, n.d.). This is done with separate projects. The focus is on the Dutch cadastre, but also other old registrations are, or will be, digitalised. The province of Friesland was the first province that was digitalised by the HISGIS project, followed shortly by the province of Groningen. The *minuutplans* were vectorised by the Fryske Akademy and placed in the Dutch coordinate system: *RD_new*. These vector maps only have the geometry of the parcels and the related section and parcel number. These attributes are later used for a join of the register. The register is digitalised by Microformat Systems and later checked by employees of the Fryske Akademy (n.d.).

The province of Drenthe was digitalised in 2016, four years later than the dataset of Groningen. The process had the same steps as the one of Groningen. However, these steps are made by different persons and institutions. The vectorisation of the parcels is done by an Indian company and the digitalisation of OATs with the help of crowdsourcing. Many volunteers helped to digitalise many records (Fryske Akademy, n.d.). More people require more guidelines and instructions to prevent errors (Bordogna et al., 2016).

The project of Drenthe was harder to complete, because of the absence of OAT and minuutplans. Parts of Beilen, Diever, Dwingelo, Emmen and Zweeloo are absent. These areas as shown as large polygons with the text unknown in the attributes. The absence of these OATs should be considered for the final analysis. Most of the records are digitalised by the crowdsourcing project. However, employees of the Fryske Akademy have done their part of the work. They checked and reviewed the records of the crowdsourced project and the vectorisation. Another step was the digitalisation of some columns, that were neglected in the crowdsourced project.

3.3 Quality assessment

In this paragraph, the content of the quality assessment is discussed. First, the choices for the metrics that are included in the assessment will be explained. This is followed by an operationalisation of these metrics in the next paragraph.

3.3.1 Quality assessment metrics for historical GIS

The literature review identified the quality metrics that are used in literature and common standards. Also, the difficulties of assessing digitised data and the experience of the users of historical GIS are highlighted. The quality of the data influences the results of the study (Knowles, 2014; Van Oort, 2006). Therefore, a quality assessing method will be developed that can be applied by historical GIS users. The method will only focus on the most important quality metrics according to literature. The articles of Goodchild & Li (2012), Bucher et al. (2016) and Knowles will be a guideline for the selection. Quality assessment is also time-consuming and since the time available for this study is limited. This further supports the choice for a limited selection.

According to Goodchild & Li (2012) and Bucher et al. (2016), the most fundamental metrics in quality assessment are:

- Positional accuracy
- Attribute accuracy
- Logical accuracy
- Completeness

Knowles (2005, 2014) adds usability as a key metrics for Historical GIS. Especially the elements about the pre-processing time and descriptions about the content of the dataset are vital in her opinion. A long pre-processing time reduces the time and resources that can be spend on the real analysis, because all this time is otherwise needed for reclassification and fixing the dataset. From Knowles point of view, manual labour should be avoided as much as possible. Manual labour is more time-consuming than automatic procedures. Also, a description of an attribute cuts time that is needed to understand the dataset. Therefore, these elements are included in the assessment as well.

In historical GIS, the attributes are used for the analyses of the study areas (Gregory, 2008; Gregory & Healey, 2007; Knowles, 2005). Therefore, these attributes should be as accurate as possible. The same goes for the completeness of the attributes. When many features do not contain attributes, they also influence the analysis. The logical accuracy is important, because gaps and overlaps can cause situations that some GIS operations are not usable. One of such operation is the check for spatial autocorrelation (Unwin, 1996).

The positional accuracy will not be performed as in other studies. The control and correction of the present cadastre to match it with the topographic situation is still a challenge according to literature (Knol & Noordman, 2003; M. H. Storm, Knotters, Schuiling, & Clement, 2015). The older the boundaries of a parcel, the more deviation it has with the boundary in the real world (Heywood et al., 2011; Longley et al., 2005). Also, the reconstruction of the parcels of the cadastre of 1832 is hard. The parcels do not have absolute coordinates in the *Rijksdriehoekstelsel* (the common coordinate system in the Netherlands), but the parcels have relative coordinates. These coordinates are calculated from a known position, on most occasions a church. Recalculation is already a challenge for a small area. The study area of this research is simply too large, and the time too limited, for such a time-consuming procedure. Also, the Fryske Akademy did a quality assessment of the digitalisation process (Fryske Akademy, n.d.). Therefore, it could be assumed that the positional accuracy is checked in the process and that the accuracy was sufficient. This is further supported by the fact that operations in historical GIS do not require a very high level of detail (Cooper & Gregory, 2011; Gregory, 2008; Knowles et al., 2008). A small distortion in the shape of the provinces or the shape of the parcel will not influence the results that much. In this research, the positional accuracy will focus on the size of the parcels.

The size of the parcels as stated in the cadastre will be compared with the size within the GIS. The comparison will check if the values in the cadastre are realistic, or that the older calculations are inaccurate. The occurrence of differences between the parcels is expected, because of the older techniques used for the measuring and calculation. Also, the digitalisation technique could fail, because of vague boundaries or corrupted paper pages. However, the comparison could still supply valuable information about the relation between OAT and the actual parcel.

The combination of the metrics of Bucher et al. (2016), Goodchild and Li (2012) and Knowles (2005, 2014) results in the following list of metrics (Table 3) that will be used to assess the Dutch digital Cadastre of 1832.

Metrics for the quality assessment of historical GIS
Attribute accuracy
Completeness
Logical accuracy
Positional accuracy
Internal usability – description of attributes
External usability – pre-process

Table 3 - Metrics for quality assessment of historical GIS

3.3.2 Operationalisation of quality assessment

The metrics of the previous section will be operationalised. The operationalisation will tell how the metrics are measured and assessed in the quality assessment, but the article of Gabriela et al. (2010) is taken as a guideline. The description will be done per metrics. The two provinces will be separately assessed on their quality. The two provinces were both a project, which were digitalised separately from each other. The results of the quality assessment will be discussed in chapter 4.

3.3.2.1 Attribute accuracy

The attribute accuracy will be measured to see if the dataset is digitalised correctly. The main assumption is that the original, analogue, source is a correct representation of the world in 1832. The analogue register can be found online as a scanned image (Rijksdienst voor het Cultureel Erfgoed, n.d.). A sample of parcels was selected from both cadastral datasets. The parcels in the sample were compared on the contents of their attributes with the attributes of their counterparts on the images.

The size of the sample is an issue. When conducting a survey, a representative size of the population should be randomly selected. If the size is sufficient, then the results can be generalised for at full population (Cowan, 1998; Unwin, 1996). However, a guideline about the size of the sample is not available for GIS quality control. Surveys can serve as example for an alternative method. There are methods that define the size of the sample for surveys (Cowan, 1998). These methods can be adopted for the quality assessment of the attribute accuracy. The following method is used to calculate the sample size:

$$ss = \frac{Z^2 * (p) * (1-p)}{c^2}$$

Where:

Z = Z value (e.g. 1.96 for 95% confidence level)
p = percentage picking a choice, expressed as decimal
(.5 used for sample size needed)
c = confidence interval, expressed as decimal
(e.g., .04 = ±4)

Figure 4 - Sample size calculation (Creative Research Systems, 2016)

A confidence level of 95% and a confidence interval of 5% are common in statistical analyses. Increasing the confidence level or reducing the confidence interval increase the sample size. However, the opposition is also true (Creative Research Systems, 2016; Unwin, 1996). Variants could be made with a higher interval and a high confidence level (Table 4).

Number of parcels (population)	Confidence level	Confidence interval	Sample Size
	Drenthe		
138840	95	5	383
138840	95	10	96
138840	99	5	661
138840	99	10	166
	Groningen		
216385	95	5	383
216385	95	10	96
216385	99	5	661
216385	99	10	166

Table 4 - Variants for the sample size calculation

The common values in statistical analyses result in a sample size of 383 parcels per province. The comparison of a total of 766 parcels is a time-consuming process. Before the actual comparison can start, the corresponding image should be searched and then the correct row should be found of the corresponding parcel. The assessment of the attribute accuracy is further hindered through the writing in the original cadastre. The handwriting is harder to read than an excel sheet filled with the same values.

To spend not too much time on the digitalisation of the attributes, there is chosen to use the smallest sample size. This means that 96 parcels should be selected. For practical issues, this is rounded to a hundred parcels per province. This means that the confidence level is slightly above the ninety-five per cent, but the difference is neglectable.

The sample of hundred parcels is selected with the ultimate suite extension of Ablebits (n.d.). This extension of Microsoft Excel offers a tool for random selections. Therefore, the attribute table is exported to a format that is compatible with Microsoft Excel. A list of the selected sample will be presented in Appendix A

In the Appendix B, the attributes that are used in the comparison are shown. In total nineteen attributes are compared. These attributes are chosen due to the fact, that some of these will be used in the analysis. Another reason is that most of these attributes should be filled. The name, place of residence, the size and the cadastral taxation are examples of these attributes. The class of the parcels (5) and the related price class are also checked, because they influence the cadastral taxation.

There are three mistakes identified that are common in digitalisation. The first is the wrongly digitalisation of the values in the original cadastre. The wrongly digitalisation is often caused due the handwriting. The wrongly digitalisation is a misinterpretation or a typing error. An example is for example that a 7 is interpreted as a 1, or a 0 as an 8. This can also occur with letters in words. The second mistake is a missing value in the digital dataset. This means that there is a value in the original cadastre, but that is absent in the digital version. The third mistake is the opposite of the second. Sometimes a value is placed in the incorrect attribute column. This results in a value that is not present in the original cadastre. The mistakes are counted in an error matrix and are assessed in the next chapter (Gabriela et al., 2010).

3.3.2.2 Completeness

The completeness will investigate the areas that are not, or partly, included in the datasets. This can be done with the selection tool. Parcels that do not have the land-use filled or have classified it as unknown are parcels that are not correctly digitised. An additional rule should be that the size of the parcels cannot be higher than zero, because parcels that do not have a size and a land-use will not be registered in the cadastre.

The taxation attribute was also named as a potential source for the completeness check, because the taxation attribute should be zero for parcels that are not digitised. However, governmental parcels will also have a zero value in the taxation attribute. So, the zero is not a unique value that shows unknown values. The location of these parcels will be later compared with the location of raised bogs. When, missing features are found near or in raised bogs areas it could have a substantial impact on the results.

3.3.2.3 Logical accuracy

In the logical accuracy, there will be a focus on the topology of the datasets. The topology will be checked with the topology checker in ArcGIS Pro. The parcels polygons must not overlap or have gaps are the main topology rules that will be checked. The errors will be counted (Gabriela et al., 2010). The topology will not be checked more in-depth, because the topology is more important when operations, like network analysis, are applied. However, for the cadastral dataset, this is less relevant.

3.3.2.4 Positional accuracy

The positional accuracy will focus, as described, on the difference between the two size attributes of a parcel. The attribute *inhoudsgrootte* have stored the value that is used in 1832. The *SizeGIS* is an attribute that is automatically calculated in a GIS. This attribute is a result of the vectorisation of the parcels. The latter value provides a detailed view on how large the parcel was if the parcel is correctly drawn. When this value is compared with the original calculated value, it could tell how correct the measurements were. However, it also tells something about the quality of cadastre and how it influences the results. When parcels are significantly smaller or larger than stated in the cadastre, other patterns could occur.

The size within the GIS is subtracted from the size of 1832. The result is stored in a new attribute '*opp_verschil*'. To interpret the results, also the relative difference is calculated. The results are shown in a Box map, a boxplot on a map, to show where the largest outliers are located and how they influence the results.

3.3.2.5 Internal usability

The internal usability is checked with the documentation and the metadata of the digital cadastre. Metadata is provided, but does it have all the necessary elements. The necessary elements are named in the ISO 19115 standard. A description and explanation of these elements can be found at the Geonovum website (2017). When all the mandatory elements (see Table 5) are included, the metadata quality can be further assessed. Guideline is the documentation of Geonovum (2016, p. 7) about assessing and improving metadata quality. This document explains when the quality of metadata is sufficient and helps to quantify the quality in this assessment. For example, the title of a dataset should be clear, without technical information and should not contain to many words (less than 75).

Metadata elements
Title of the dataset
Date of the source
Date of publication
Summary
Status
Subject
Responsible organisation
Contact information of the organisation
Name of contact person
The profession of contact person
Legal restrictions
Restrictions in use
Coordinate system

Table 5 - Mandatory metadata elements (Geonovum, 2017)

Another aspect that will be checked and described in the results of the internal usability is about the documentation of the attributes. The attributes do not have clear names and therefore it is harder to identify to which aspect of the original cadastre it is related. Also, some attributes are added by the Fryske Akademy. These attributes are the results of reclassifications and editing. The documentation will be assessed to see if the attributes are described and how they are created.

3.3.2.6 External usability

The external usability involves the pre-process of the dataset, before it can be used in a GIS analysis. The assessment of this process tries to take the experience of the users into account. Some tasks are harder for non-GIS specialist than for people that are more familiar with GIS. (Cowan, 1998; Hennig & Belgui, 2011). The goal of a pre-process is a smoother and clearer analysis of the cadastral situation in 1832. Due to proper pre-processing, the results are much easier to interpret. Important to note is that the external usability focusses on fixing and editing the cadastral dataset. And thus, make the original dataset available for analysis. The computation of additional attributes, which will be done for the final analysis, is neglected for the external usability. The process of these additional attributes is described in another paragraph.

There are three main steps identified for the pre-processing of the Dutch cadastre of 1832. These steps will be described and graded. The grade reflects the level of experience that is needed to complete the task and the time needed for this task. The grade should tell the potential users of the effort that should be put in. For the pre-processing of the attributes, the software of FME is used. FME is software created by Safe Software that can convert many data types from A to B. It is also frequently used by GIS companies and municipalities to automate workflows and transform the data (Safe Software, 2019). The main advantage is that the software provides an overview of the workflow and that it does not create unnecessary data until you want to have a result.

The first step in pre-processing is to standardise the units of the taxation class. The taxation attribute is not uniform filled. The attribute in Drenthe is in cents, where the data in Groningen is in whole Guldens. The data of Drenthe is divided by hundred to convert the cents to Guldens.

The second step is the reclassification of some attributes. The profession and the land-use attributes are two examples that could be reclassed. Without reclassification, the results are much harder to analyse and to interpret, because of the many different professions and land-uses. In this study, the focus is on the land use of a parcel, and not on the profession of people. Therefore, only the land-use attribute is reclassed to broader values. This is further stimulated, because the author could profit from earlier work of HISGIS. For the Drenthe dataset, HISGIS has already made a reclassed land-use attribute, which narrowed the land-uses to seventeen classes. Where there were two overlaps in the reclassed attributes, for example, '*dijken*' and '*dijk*' are the same, because these two were differently written, the GIS software sees these two as two separate classes. The same goes for '*onbekend*' and the empty values. Empty does also mean unknown, so these could also be added together. The original and reclassed attributes were the base for the reclassification process.

The reclassification process is shown in Appendix C. The scripts are divided into multiple colours. Only the same classes are not represented in the scripts and are done manually after the process. The grey coloured script is used to get the classification of the dataset of Drenthe. The workflow contains multiple actions. The attribute manager transforms all the values of the land uses attributes into lowercase text. The software is susceptible for differences in capital- and lowercase letters. By eliminating the difference, more matches could be made. The next step was the statistics calculator to group all the unique values in the land-use original and display the reclassed variant in a second column. The next transformers help to create an excel file. This file has two columns; one with the original land use and the second column has the reclassed value. This excel file is the reclassification scheme.

The application of the scheme is performed in the turquoise coloured script. It starts again with an attributemanager for converting text to lowercase letters. The AttributeValueMapper, applies he reclassification scheme on the dataset of Groningen. It looks for all the land-uses in the Groningen dataset, if there is a value that corresponds with a value in the classification scheme. When a match is found, the reclassed variant is copied to a new create attribute. The result is that the Groningen dataset also has an original and a reclassed land-use dataset. Unfortunately, not all parcels in Groningen have an original land use, that can be matched with the land use of Drenthe. The purple script filters the unassigned parcels and groups the unique value and writes them to Microsoft Excel. The classification for these unique values is done manually. In the end, the list is copied to the automatically generated classification. The last step is to rerun the turquoise script and check if some parcels are still unmatched. The result is that all parcels have an original and reclassed land-use.

The third step in the pre-processing cycle is to merge the Drenthe and Groningen dataset. The attributemanager of FME is used to match the attributes names of both datasets. In this occasion, the names of Drenthe are unchanged, and the ones of Groningen were edited to match the Drenthe attribute names. This was quite arbitrary and could also be done the other way around. The attributeKeeper transformer is used to make sure that only necessary attributes are kept. At last, the two datasets are combined in one writer. A writer combines the two datasets as one layer of a geodatabase.

3.4 Distinction of soil types within GIS

In this paragraph, the distinction of raised bogs in a GIS is addressed. A dataset with a single raised bog category that covers the year 1832 is not available. Therefore, alternative solutions must be found for the raised bog distinction. The main challenge for the raised bogs distinction is the availability of multiple sources holding data related to bogs. However, most of the sources only show the present situation, where this research focus on raised bogs in a historical context. Another challenge about the datasets is that raised bogs are often not represented as a single category. This means that always other land covers are selected and analysed as well. However, this paragraph will produce three different methodologies to identify raised bogs.

Theo Spek (2004) offers the first methodology for raised bogs distinction. Worst (2012) and Grift (2015) apply this methodology in their studies. The methodology of Spek is based on the soil map of the Netherlands (Alterra, 2006). Spek supplies a list of soil classes, that contain traces of disappeared raised bogs. This list plus the soils of former peat colonies are all the raised bogs soils in Drenthe. The soils of the peat colonies are added, because of the presence of sphagnum moss bogs in almost all peat colonial soils (Spek, 2004). However, how the peat colonies are found within the soil map is not well described. Therefore, the article of Pleijter (2004) is used. The article of Pleijter describes that the term: *Veenkoloniaal*, is used for soils within former peat colonies. Therefore, all the classes of the soil map of the Netherlands that contain this term are added to the list of Spek (see Appendix D). When the result (Appendix E) is compared with the total reconstruction of Spek (Appendix F), there can be seen that the same areas are represented. The greatest differences occur due to the generalisations that Spek made based on other sources, like soil drillings and names of areas.

The second methodology is based on the BOFEK map (Alterra, 2018). BOFEK, which is sort for '*Bodemfysische eenhedenkaart'*, is less familiar in the literature compared with the soil map. The BOFEK map is used in the research of De Vries et al. (2008). In this research, the disappearance of bogs and the impact on the environment is analysed. However, De Vries et al. does not supply a method for raised bogs distinction. The method to use the BOFEK map should be created before it can be used in this research. The classes of the BOFEK have a number. These numbers correspond to a unique soil class. The numbers that are potentially (former) raised bogs are selected (Appendix G). The BOFEK map has five main categories. The numbers in the hundreds are related to bogs. The bogs are separated into ten different classes. The bogs classes have various classes that are potentially raised bogs. Soils that do not have clay within them are potentially raised bogs. Clay indicates the presence of water bodies and raised bogs were mostly on the higher grounds (F. De Vries et al., 2008; Jong, 2009; Spek, 2004). In chapter 2, there is described that raised bogs are oligotrophic, which is also a class in the BOFEK map. Also, soils with the terms of peat colonial (*veenkoloniaal*) (Pleijter, 2004), are added to the selection of potentially raised bogs. For the classes that fall within *moerige gronden*, only two classes are identified that potentially are raised bogs areas. After the peat exploitation, the soils were covered with sand (Casparie & Molema, 2015; Worst, 2012). The classes 205 and 206 contain soils that could be a leftover of peat exploitation. The final BOFEK map is shown in Appendix H.

With the help of the methodologies of Spek and Pleijter and the BOFEK map, two layers with raised bogs areas are created. Both methodologies have some flaws. The method of Spek is only accepted for the province of Drenthe, where the study area of this research also covers the province of Groningen. The method with the BOFEK map is not widely used and therefore unknown how well it performs. Both methodologies identify other areas as raised bogs. However, the similarities could help to make more certain which areas are actually raised bogs. Areas that are both covered in the method of Spek and Pleijter as the BOFEK method are certain raised bogs. Therefore, a combination methodology, where the previous two overlap, is also used in the analysis.

The last step is to assign which parcels are covered by the different outputs of the methodologies. For these three methods, columns are added to the cadastre of 1832. These columns are later filled with a zero or a one. The zero means that the parcel lay not within a raised bog area, the one is for a parcel that does lay within a raised bog area. The select by location operation is used to select which parcels lay within a raised bog area. The select by location operation has multiple options for the selection tool. The intersect method assigns all the values that intersect with the dataset, even when it is a small fraction of the parcel. Other options, like the within tool, only select parcels that are fully covered by the other layer. This method is therefore, stricter than the intersect tool. This is a safer method, because it only assigns a value when it is more certain that they are raised bogs. Another possibility is to use the contain option. This option only selects parcels that overlap a raised bogs area and does completely the opposite of the within tool. Three options will be used to compare the findings of the analysis and show the impact of the choice. Parcels that are highlighted as with the contain/within tool, are also present in the intersect method. However, parcels that are selected with the contain tool cannot be present in the within tool and vice versa.

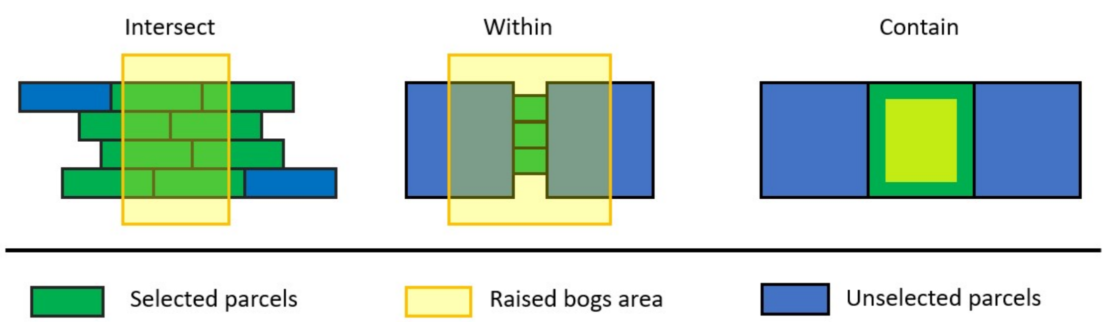


Figure 5 - Schematic overview of the intersect, within and contain options in the select by location tool

The above results in three methodologies that are used. The methodology of Spek and Pleijter, the BOFEK and a combination. The combination offers more certainty when an area is identified as raised bog. These methodologies are assigned to parcels with the select by location tool. This tool offers multiple options, where three are selected. One broader method (intersect) and to different stricter method (contain and within). This results in a total of nine different datasets (three times three).

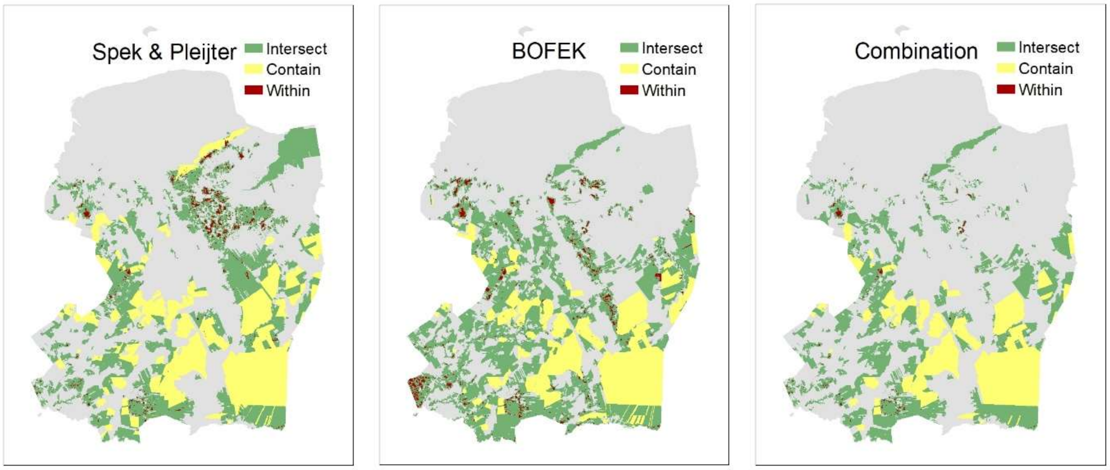


Figure 6 - The areas selected with the three selection methods and the three options of the select by locations tool

3.5 The study of raised bogs areas with the cadastre of 1832

The analysis of the Dutch the cadastre for the raised bogs areas will focus on one specific use-case. The case is chosen due to the lack of studies that specifically focus on this theme. Current studies neglect the smaller raised bogs areas and primary focus on the peat colonies. This assumes that the smaller areas have followed the same dynamics as these peat colonies. However, the theses of Spek (Spek, 2004), Worst (2012), Reeskamp (2013) and Grift (2015) show that each region has its unique characteristics. A few of these characteristics can be derived from the cadastre. Also, knowing the regional similarities or differences could help with the interpretation of other studies. However, before these cases are analysed, the analysis will focus on the selection methodologies. These will also be analysed to see the impact of the choice of a specific selection methodology.

The analysis will be performed with multiple attributes of the Dutch cadastre of 1832. Chosen is to a primary focus on the attributes related to the parcel. The names, profession and residence place of the owner will not be continuously used in this analysis. The attributes that will be often used are the reclassed land-use attribute, the taxation and the size of the parcel. This are the same attributes as in the study of Knol and Noordman (2003). The land-use is reclassed in an earlier stage of the study and results in a nominal scale. The other two attributes are ratio scale. The size of a parcel is stored in two attributes. One has the value as initially stated in the cadastre (*inhoudsgrootte*), the second stores the size of the parcel within a GIS. This second attribute is called '*sizeGIS'* in the remaining of the report. The taxation is initially stated as an amount per parcel, however, to compare the taxation for larger areas two additional attributes are created. These two attributes store the taxation per hectare. One is calculated by dividing the taxation by the *inhoudsgrootte* and the other is calculated with the help of '*sizeGIS'*.

The analysis will involve the summary statistics tool. The summary statistics tool can summary various statistics of an attribute, like the mean, the maximum value, and the sum. The summary statistics tool can be used, when a selection is active. This means that only the selected parcels will be analysed. The summary statistics tool is also enormously powerful in combining different attributes. An example of an operation is to summarise the average size of a parcel of all the unique land uses for a specified selection methodology.

Attributes	Scale
Land use (*Soort Eigendom gemapped*)	Nominal
Taxation per parcel (*Kadastraal inkomen*)	Ratio
Taxation per hectare (*inhoudsgrootte*)	Ratio
Taxation per hectare (*sizeGIS*)	Ratio
Size *(inhoudsgrootte*)	Ratio
Size (*sizeGIS*)	Ratio

Table 6 - The attributes for the analysis and the corresponding scale

The use-case for this research, as explained, focuses on the concept of small and large raised bogs systems. For the distinction, For the distinction of large and small bogs, the dataset 'HISTLAND' is used (Rijksdienst voor Cultureel Erfgoed & Ministerie van Onderwijs, n.d.). This dataset holds information about the location of peat colonies. The main peat colonies were all located on a system of raised bogs in Drenthe and Groningen (Casparie, 1993; Gerding, 1995). So, the larger bogs are within the peat colonies, because they could be more efficiently be exploited. The smaller bogs are outside these areas in this research. A downside of this distinction is that the smaller bogs systems are not always small bogs. The remaining bogs in Bargerveen and Fochteloërveen are also part of the smaller raised bogs systems. The analysis will compare the five main peat colonies, that are part of the HISTLAND dataset (Figure 7). The areas that are outside the peat colonies are grouped and compared with these five areas. Differences between the peat colonies can be seen, as well as the differences between large and small areas.

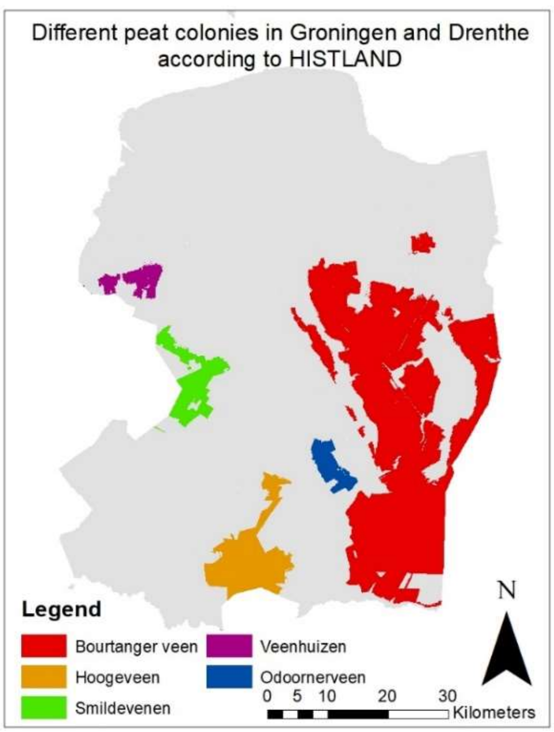


Figure 7 - Different peat colonies in study area according to HISTLAND dataset (larger version in Appendix I)

3.6 Overview

Below, a schematic overview of the research is presented (Figure 8). All the steps that are described in the methodology are shown. Three parts can be identified. The quality assessment (blue) will be done during and right after the pre-processing. The pre-processing is specific for the Dutch digital cadastre. This means the standardisation of units, the reclassification, and the merge of the datasets. The time needed for the pre-process will be used in the quality assessment. The red part is the actual analysis of the Dutch cadastre. However, the green track makes it case-specific for raised bogs, without the green track the full cadastre is analysed. In other studies, the content of the green track will differ, because of another use-case and goals. The green track contains for this study the distinction of raised bogs areas and the distinction of large and small bog systems.

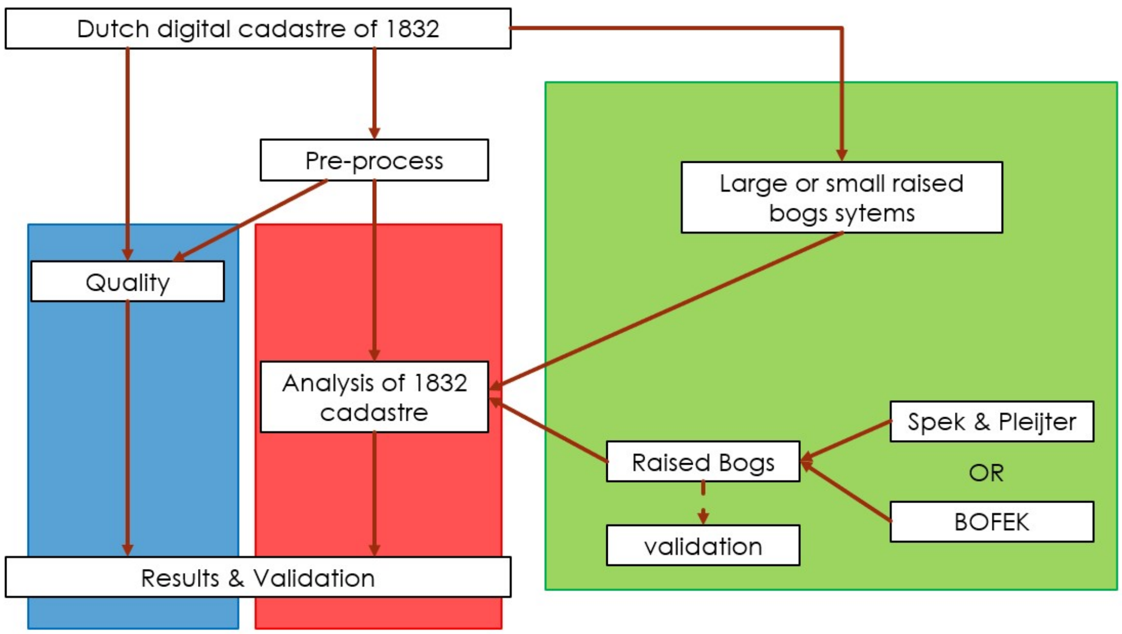


Figure 8 - Schematic overview of the methodology

4. Quality assessment results

This chapter will focus on the results of the quality assessment. The results are described per distinguished element. Within an element both datasets (Drenthe and Groningen) are described. The final paragraph will conclude the results and tries to quantify the overall quality of the digital cadastre of 1832 that is provided by HISGIS project (Fryske Akademy, n.d.).

4.1 Attribute accuracy

The attribute accuracy is executed to assess nineteen different attributes. The attributes and the results are shown in Appendix B. The objectID, parcel number and the corresponding OAT of the sample are also provided in the appendix A. The samples contain a hundred features per dataset. Multiply this with the nineteen attributes and this results in 1900 elements that should be checked. There are three potential mistakes identified in the operationalisation that affect the accuracy of the attributes.

For the province of Drenthe, a total of 51 errors are observed. This means that 2,68% of all the values contain an error. The error that is the most common is the wrong digitalisation of the values — seventy-five per cent of all the errors caused by this. The feature with ObjectID number 78507 contains four errors in the attributes related to the owner of the parcel. The names contain typing errors, and the profession and place of residence are lacking. These are two of the three missing values in the sample of Drenthe. The third mistake, incorrectly filled attribute, is observed ten times. Eight of these mistakes are observed in the 'Tarief1' class. These fields should be empty, because the second and third classes of these features were correctly filled. The attributes that have the most errors are the profession and the cadastral taxation attribute.

One remark should be made related to the features that are used in the sample of Drenthe. Four features are replaced, because of an incorrect OAT value. All the four features were related to a municipality. Maybe, the municipalities were differently processed in the Drenthe project. The four features were replaced, with four other randomly chosen features.

The dataset of Groningen does not contain added attributes. Also, the attributes that are checked looked clearer. Maybe the dataset of Groningen had an additional quality check before it was distributed to the author? In total seven mistakes are found. All seven are a misinterpretation error. Four of them are made in the prices of features. Two times in the price classes, which also results in an error in the cadastral taxation attribute, because the sum of the values in the price classes should be the cadastral taxation, an error in the price classes influences the taxation attribute. The misinterpretation was not of a high order — only a cent difference between the value in the digital version and the original cadastre. The impact of the difference in cents in the taxation class is limited. The other three errors are made in the names and profession of the owner. The impact in the names is not problematic, because the names of people are hard to use in a GIS analysis. An error in the profession could have a higher impact, because these attributes could be reclassed and be used in the analysis.

	Drenthe		Groningen		Total	
	Abs	*%*	*Abs*	*%*	*Abs*	*%*
Misinterpretation	38	74,5	7	100	45	77,6
Missing value	3	5,9	0		3	5,2
Incorrectly filled	10	19,6	0		10	17,2
		100		100		100
Total dataset	51	2,68	7	0,37	58	1,53

Table 7 - Attribute accuracy error matrix

To conclude, the attribute accuracy differs between Drenthe and Groningen. The accuracy of Groningen is particularly good, which only 0,37% of the fields that are incorrect. The dataset of Drenthe contains more errors, 2,68% of the fields contain an error. When the sample size is increased, the percentage could be much higher. This is related to the high confidence interval. The results that could be 10% higher than originally is found in this assessment. This means that the percentage of the errors could drop below the ninety per cent. A value of less than ninety per cent is lower than the norm that is used by the Dutch government for their Basic Registration Topography (M. Storm et al., 2018). However, this dataset is not a key dataset of the government. Based on the results of the attribute accuracy assessment, there could be argued that the attribute accuracy of these datasets is quite good. There could be some improvement in the Drenthe dataset. A larger sample can provide more accurate results about the true percentage of errors that occur in this dataset.

4.2 Completeness

The completeness of the dataset is important to assess if all features that should be included are indeed included in this dataset. The attributes, that will play a key role in the final analysis, will be reviewed. The first is done with the help of a selection of all the parcels that have empty or unknown values in the land use class. The results of these could be compared with another layer provided by the Fryske Akademy (Fryske Akademy, n.d.). This layer contains all the parcels that were absent in the 1832 cadastre. In a previous chapter, there is highlighted that of several municipalities the OATs were absent. The data layer does not have any attribute and is only a visual representation of the parcels that were missing.

Based on the selection 315 parcels do not have a land use registered and have a size of zero according to the digital cadastre. A quick scan of the surrounding values shows that they are either zero, empty or unknown. The layer of 1862 does not completely cover the missing parcels in the 1832 dataset. The missing parcels in the 1832 dataset cover a larger area (Figure 9; Table 8). The total amount of square kilometres lays around the 312. The area around Emmen is the largest that is missing. But also, the area of Zweeloo and the area around Hijken is not present in digital cadastre. The area around Hijken, which is part of the municipality of Beilen, should be present. The OATs are available for these areas, as well as the mapping pages. The same is the case for the area of Zweeloo. The area of Emmen is only partly available in the *Beeldbank* (Rijksdienst voor het Cultureel Erfgoed, n.d.). Only sections A to D are available. Based on this information, we can confirm that parts of Emmen are not available in the cadastre and cannot be mapped therefore in the digital version. However, it is strange that the part that is available is not mapped as well. Therefore, we could argue that the dataset of Drenthe is not a complete representation of the analogue cadastre of Drenthe.

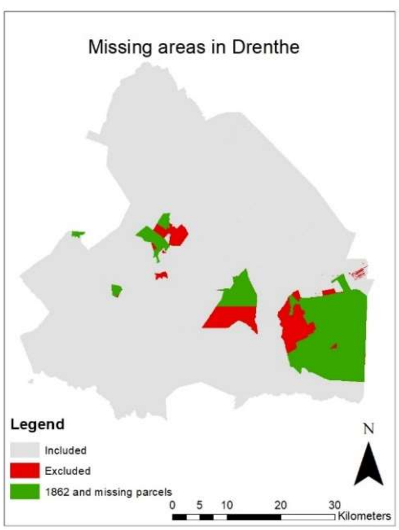


Figure 9 - Missing areas in the province of Drenthe

Name	Number of parcels	Size of the area (km^2)
Emmen en Westenesch	72	2
Emmen	1	213
Hijken	4	29
Noordlhe (dwingeloo)	1	3
Roswinkel	224	2
Weerdinge	12	0.052
Zweeloo	1	63
Total	315	312.052

Table 8 - Number of parcels and size of missing areas

In the previous paragraph, there is identified that the dataset of Groningen looked more accurate than the Drenthe dataset. Also, in terms of completeness, the dataset of Groningen contains less error. Only 32 parcels, with a total size of 400 square meters, do have an unknown value for the land use class. Also, the surrounding values like the names and the size are unknown or zero. Which shows that these parcels are not complete. When the OAT reference is used to search the original scans, it becomes clear why these parcels do not have any values. The scans only show the right page, and therefore not the whole information can be used. The information about the cadastral taxation can also not be linked to a parcel, because the parcel number is also displayed on the absent left-page. Other parcels are all included in the dataset of Groningen. There can be concluded that the completeness of Groningen is much better than the completeness of Drenthe.

4.3 Logical accuracy

The logical accuracy is checked with the help of the topology checker of ArcGIS pro. Main goal is to observe the gaps and overlaps between parcels. These gaps and overlaps should be eliminated to do more advanced spatial analysis, like a network analysis or observing clusters. The topology checker identifies how many errors occur with the current topology. The topology rules focus on that features in own dataset cannot overlap and cannot have gaps. The dataset of Groningen contains many more errors in comparison with Drenthe (Table 9). Fixing the topology errors is required to enable the more advanced analysis. The easiest way to do that is to assign a standard fix. For example, merge the overlap to the largest parcel and remove it by the smallest. Other fixes are also available, but these are not in the scope of this project.

Dataset	**Rule**	**Number of errors**	**Relative amount**
Drenthe	Parcels must not have gaps	330	0.24%
	Parcels must not overlap	1133	0.81%
Groningen	Parcels must not have gaps	1630	0.75%
	Parcels must not overlap	17056	7.88%

Table 9 - Logical accuracy matrix

4.4 Positional accuracy

For the positional accuracy, the difference in size of the parcels is investigated. As told earlier, the datasets have two attributes that store the size of a parcel. One is the size as initially calculated in the cadastre of 1832. The other one stores the area of the polygon shape that represents the parcel within the GIS. First, the total size of the study area is calculated. According to the cadastre Groningen and Drenthe (excluding missing parcel) is 6164 square kilometres. The size within the GIS is only 5017 (without the missing parcels). This is a significant difference and can influence the results of the analysis of raised bogs areas. Therefore, we could argue that the significant difference reduces the quality of the dataset. However, the differences do not explain why these differences occur. Multiple reasons could be the cause of the difference. Minor differences between the size in 1832 and the size of the polygon is expected. The area in 1832 is calculated with less advanced and precise techniques than we could do nowadays (Longley et al., 2005). Therefore, the values of the two attributes will not be identical. However, it could also be a digitalisation error, in both, the register as the vectorisation of the parcel.

To spatially analyse where the differences also occur the relative difference is calculated. These differences are displayed in a Box Map, a visual representation of a boxplot. The Box Map shows the relative differences in six classes. The first and last categories are the outliers. The lower outliers are the outliers were the *inhoudsgrootte* is higher than the *sizeGIS*. The upper outliers are the opposite. In this case, the lower outliers start where the difference is more than -17.15%. The upper outliers start with a difference of 17.36% and more.

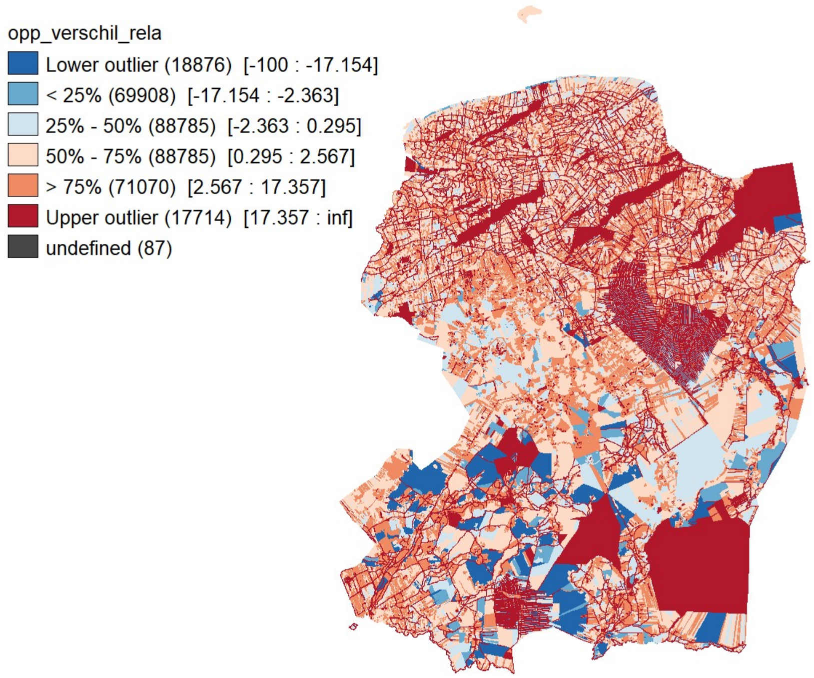


Figure 10 - Box Map of the study area with the relative differences in size between the size stated in the register (from 1832) and the size of the vectorised parcel

Based on the map (Figure 10), Groningen looks to have more upper outliers than the province of Drenthe. Thus, the vectorised parcels are larger than stated in the cadastre of 1832. Significantly, most of the outliers are part of networks (water, roads, and dykes) in both Drenthe and Groningen. This is even better to see on figure 11, where you clearly see the networks with the surrounding parcels. The other parcels also have differences, but these are often less significant. In Drenthe, the missing parcels are also showed as an upper outlier. The reason for this is that the *inhoudsgrootte* is zero for these areas. This leads to an infinity relative difference in this map and therefore upper outliers. These upper outliers in Drenthe lay in the large raised bogs areas. Therefore, the analysis will be less reliable.

The lower outliers are the parcels where the size of 1832 is larger than the size of the vectorised parcel. Some of these lower outliers are also outliers, especially in the north of Groningen. However, Drenthe has the largest parcels that are part of the lower outlier category. For these parcels, the other attributes do not have a value in common.

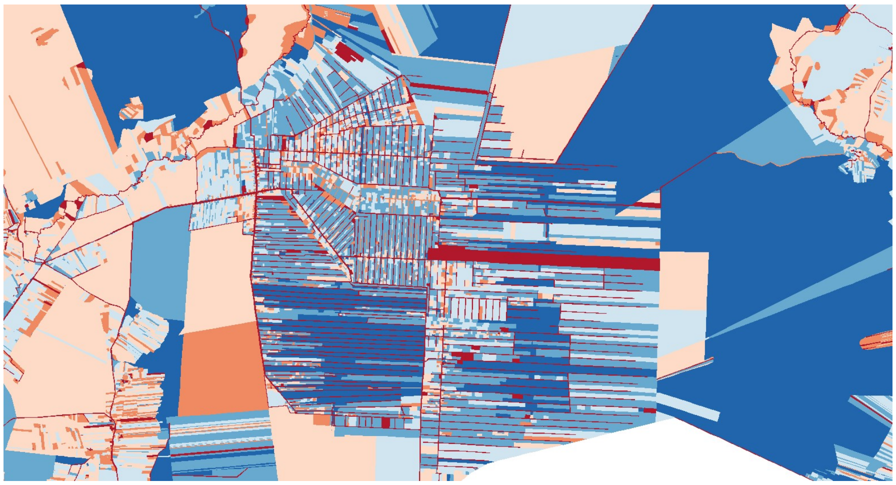


Figure 11 - More detailed impression of the Box Map, for the legend see figure 10.

In total 36590 parcels are identified as outliers, which means that all these parcels a seventeen per cent or more difference between the original and the size within the GIS. Due to this high amount, both attributes will be used in the remaining of the analysis. This reduces the possibility of incorrect conclusions, because of incorrect sizes in the original cadastre. It is also an opportunity to observe if different patterns occur when different size attributes are used. The usage of two attributes in the analysis means, for example, that the taxation per hectare is calculated both with the *inhoudsgrootte* attribute as with the size within the GIS.

4.5 Internal usability

The metadata will be reviewed to assess the internal usability of the datasets. First, the metadata is checked on their contents. Does the metadata contain all the mandatory elements according to the Geonovum (2017) guidelines? The next step is to assess the quality of these elements. Guideline for this is the Geonovum quality report of 2016 (Geonovum, 2016). This document provides some guidelines when a title is good and when a title is too long or too complicated. The quality can be good, neutral, or bad.

Both datasets use the same format for metadata description. Therefore, one table (Table 10) is created to assess the format. All elements were present in the format of metadata description, except the profession of the contact person. Further, no outstanding observations are made regarding the metadata elements. The quality is overall very well, with only the contact information that scored a neutral and a low score. The contact information relates only to a specific person at the Fryske Akademy, and not to the Fryske Akademy itself, which is the criteria used by Geonovum (2016). The profession of the contact person is also not mentioned in the metadata. The profession can provide information about the level of knowledge of that person, which changes the questions that could be asked regarding the dataset.

Metadata element	Present in metadata	Quality
Title of the dataset	Yes	Good
Date of the source	Yes	Good
Date of publication	Yes	Good
Summary	Yes	Good
Status	Yes	Good
Subject	Yes	Good
Responsible organisation	Yes	Good
Contact information of the organisation	Yes	neutral
Name of contact person	Yes	Good
The profession of contact person	No	Low
Legal restrictions	Yes	Good
Restrictions in use	Yes	Good
Coordinate system	Yes	Good

Table 10 - Metadata quality assessment matrix

The last comment is made regarding the contents of the Drenthe dataset. The Drenthe dataset does not hold only the attributes of the original cadastre, but also some additional attributes. The process of how these attributes are derived is not described in the metadata. This makes it difficult to implement these attributes in your own analysis easily. People that are less familiar with GIS and/or cadastre data cannot implement these attributes in their research. In the analysis of this project, they are mostly neglected. Only the processed land use class will be used for deriving the classification scheme (see next paragraph).

4.6 External usability

The external usability will focus on the pre-processing time needed to make the dataset suitable for your purpose. For this project, the purpose of the dataset is to analyse raised bogs areas. In the operationalisation chapter, the three necessary steps are described. The first step was the standardisation of the units in the price classes and the taxation class. The units of Drenthe are in cents, and the units of Groningen are in Guldens. This is corrected by divide the units of Drenthe with 100. This is a simple operation, even for less experienced GIS-users.

The second step is the reclassification of the land-use class. As explained earlier, the dataset of Drenthe holds already a reclassed land use class. The task is to derive the classification scheme that is used. The process was done with help of FME. The classification is derived from Drenthe and then applied on Groningen. The unmatched land-use are manually inserted in the classification scheme. After the completion of that step the classification is again applied on the dataset of Groningen. FME provides helpful tools for the process, However it requires expertise with the FME software. The software is not freely available and is not often part of GIS courses at universities. Therefore, it is unlikely that none GIS-experts have enough ability and affinity with the software to do this task.

The last step is a simple operation that needs little time in comparison with the reclassification. The merge results in one dataset that covers both Groningen and Drenthe. During the merging process, all the unnecessary attributes are dropped.

4.7 Overall assessment

The function of this paragraph is summarizing the earlier paragraphs and to try to quantify the found results. This quantification is needed to estimate how the quality affects the results of the final analysis of this project. The usability of the datasets is quite well. The metadata is complete and has a high quality, with only one element that did not meet the norm of Geonovum. The additional attributes and its processes are not described, which makes it harder to implement them. However, this is a downside that is not of vital importance, because they were not originally part of the cadastre of 1832. The external usability is also quite high. The standardisation and the merging are easy processes; however, the reclassification is a time-consuming job especially when it is done manually.

The logical accuracy assessment found many errors for both provinces, but Groningen had more errors than Drenthe. This makes it harder to use these attributes in more advanced analysis without modification (Sırrı Mara, Hadi Mara, Aktu, Emin Mara, & Yildiz, 2010). The positional accuracy shows many differences in size between the original size in the cadastre and the size within the GIS. Some of the outliers could be identify as roads, waterways, and dykes. Others, like the missing parcels, are known raised bogs areas. Some lower outliers occur also in parcels that do not have a network function. These lower outliers are mainly concentrated in Drenthe.

Regarding the completeness and attribute accuracy, the dataset of Drenthe scores less in quality than the Groningen dataset. The Groningen dataset holds almost all parcels, just one OAT page is not digitalised. Also, the attribute accuracy is high with only seven errors and only in attributes that are not further used in the analysis. Drenthe, however, misses more than 312 square kilometres in parcels. Most of the parcels are (now) present in the digital library (Rijksdienst voor het Cultureel Erfgoed, n.d.). The attribute accuracy is also lower than Groningen, but with 2.68% of the fields containing an error still quite well. However, these results are influenced by the small sample size. The actual number of errors could be higher for both provinces. To check this, a larger sample size should be used next time.

In this research, the completeness, attribute accuracy and positional accuracy are more important than the logical accuracy. The focus is on study of the attributes of the Dutch cadastre of 1832. A good logical accuracy enables extra possibilities, but do not hinder the scope of the project. Overall, the quality of both is quite well, but there could be argued that the dataset of Groningen has a slightly higher quality compared with the Drenthe dataset. Most elements do not differ much between the two datasets, but the low completeness of the province of Drenthe reduces the quality of the dataset.

5. Results of the study of raised bogs

After the quality assessment, the focus will shift to the study of raised bogs. To study the raised bogs, various methodologies are operationalised and reviewed. The results will be described first in this chapter. The next step is to study raised bogs to attain knowledge about their characteristics in the Dutch cadastre of 1832.

There are multiple selection methodologies, and therefore they are given a code that serves an identification. The code is not complicated and all and represents the name of the methodology as the first part. These methodologies are BOFEK, Spek (& Pleijter) and Comb (short for combination). The second part is the option that is used in the select by location tool to assign parcels as raised bogs. The potential options for the second parts are Intersect, within and contain. For example, the code Spek_Contain means that the methodology of Spek & Pleijter is used for the raised bogs areas and that the contain option is used to assign parcels as raised bogs. Another example is the Comb_within. This means that the combination of BOFEK and Spek & Pleijter is used, thus the parcels that overlap in these methodologies, and that they are assigned to a parcel with the within option.

Other names that are often used in this chapter are *inhoudsgrootte*, which is the size that is stated in the register of 1832. Others are *sizeGIS* or *size within GIS*, which indicate that the size is calculated of the vectorised parcels within the GIS.

5.1 Analysis of selection methodologies

This paragraph will compare the statistics of the nine possible selection methods. There will be an explicit focus on the number of parcels, the size of these parcels and the taxation that is paid in these areas. These attributes are the same as in the study of Knol and Noordman (2003). The size of the parcels is calculated both for the size that is stated within the cadastre, as for the size that is calculated within the GIS as a result of the vectorisation of parcels. In the previous chapter, it is observed, that there are some differences between these two attributes. A comparison of the results for both attributes could help with the explanation of the results, that will be found during this project.

In figure 6, the spatial distribution of the parcels for all the selection methods is shown. For these methodologies, the corresponding number of parcels is shown in table 11. The table groups the methodologies on the select by location option. Major differences occur between the options. The contain tool only select 56 to 119 parcels, but the intersect option selects almost 48.000 parcels. The differences between the three methodologies are less. The Spek and BOFEK select more or less the same number of parcels, where the gap in the intersect option is the largest with more than five thousand parcels as a difference. However, the combination method is much lower, than the other two. This suggests that the Spek and BOFEK methodology do not identify the same locations as raised bogs. For the contain option only 23.30% of the parcels are raised bogs in both methods, but in the within option, the overlap is even less, with only 13.18% of the parcels that overlap.

Selection method	**Number of parcels**
Bofek_Contain	117
Comb_Contain	56
Spek_Contain	119
Bofek_Intersect	47946
Comb_Intersect	18384
Spek_Intersect	42862
Bofek_Within	11353
Comb_Within	2737
Spek_Within	9407

Table 11 - Number of parcels for the selection methods

The distribution of the number of parcels is not the same for the size of the areas that are selected as raised bogs (Table 12). First, *inhoudsgrootte* is compared with each other. The total area of contain methodologies, with the lowest number of parcels, is far larger than the methodologies that used the within tool. The area of the within tool is tiny compared to the others. Also, the maximum and average size of a parcel is less than with the other selection tools. The maximum size of a parcel is 157 hectares with the within option and the average parcel size is just under the one hectare. The intersect options have the largest total size of the area, but because of the large number of parcels, the average size is not meeting the contain values. With an average size of 4.47 to 7.97 hectares, the sizes are higher than the within options but far less than the contain options. The maximum size of a parcel is the same for contain and intersect, which means that the same parcel is selected in both methods. The average size of the contain methods differs between the three methodologies. The BOFEK has the smallest average size, with almost 100 hectares less than the methodology of Spek. The combined method has the smallest total size, but the highest average due to the small number of parcels in this method.

The other size attribute provides a similar pattern as the *inhoudsgrootte*. The average and total area of the contain option are larger than the *inhoudsgrootte*, but the BOFEK method has still the smallest average. The methodology of Spek has the largest total area of the three methods. The maximum value is far larger in size within a GIS than in the *inhoudsgrootte*. The intersect and within options have higher values in the *inhoudsgrootte* than the size within the GIS for the average and total area. However, the intersect option with the Spek methodology has a larger average and total area.

Size for *inhoudsgrootte* (in hectare)			
Selection method	**Maximum**	**Average**	**Total area**
Bofek_Contain	3048	381.58	44645
Comb_Contain	3048	613.70	34367
Spek_Contain	3048	482.59	57428
Bofek_Intersect	3048	4.47	214244
Comb_Intersect	3048	7.97	146437
Spek_Intersect	3048	4.49	192527
Bofek_Within	157	0.65	7406
Comb_Within	157	0.66	1801
Spek_Within	157	0.66	6209

Table 12 - Size in inhoudsgrootte, in hectare, for the selection methodologies

Size within GIS (in hectare)			
Selection method	**Maximum**	**Average**	**Total area**
Bofek_Contain	21374	574.51	67217
Comb_Contain	21374	1044.84	58511
Spek_Contain	21374	684.95	81510
Bofek_Intersect	21374	4.27	204879
Comb_Intersect	21374	7.91	145382
Spek_Intersect	21374	4.58	196513
Bofek_Within	16	0.56	6384
Comb_Within	16	0.46	1263
Spek_Within	16	0.50	4713

Table 13 - Size calculated in GIS for the selection methodologies

Difference *inhoudsgrootte* versus Size within GIS			
Selection method	**Maximum**	**Average**	**Total area**
Bofek_Contain	-18326	-192.93	-22572
Comb_Contain	-18326	-431.14	-24144
Spek_Contain	-18326	-202.36	-24082
Bofek_Intersect	-18326	0.2	9365
Comb_Intersect	-18326	0.06	1055
Spek_Intersect	-18326	-0.09	-3986
Bofek_Within	141	0.09	1022
Comb_Within	141	0.2	538
Spek_Within	141	0.16	1496

Table 14 - Differences in size for the selection methodologies

Now, we will focus on the taxation differences between the nine selection methods. Therefore, the average and maximum taxation will be compared. Not only the taxation per parcel but also the taxation per hectare is calculated and compared. For the taxation per hectare, both size attributes are used. The results are shown in table 15, and the results show that the contain options have the highest taxation per parcel. The values are far higher than the values of the intersect and within. The methodologies using the intersect option have an average that is almost equal for all methods. The within options have even smaller averages per parcel. However, when the taxation per hectare is calculated a different pattern can be seen. Now, the within options have the highest average and the contain have the lowest average per parcel. The amount of taxation is higher when the size within GIS is used in the calculation. This indicates that the parcels are smaller when the parcels are vectorised compared with the size stated in the original cadastre register. The maximum taxation per hectare is quite high in the size within GIS category. This indicates that there are some very small parcels that have a relatively high taxation value.

Taxation per selection methodology (in Guldens)						
	Taxation per parcel		**Taxation per hectare (*inhoudsgrootte*)**		**Taxation per hectare (Size within GIS)**	
Selection method	**Average**	**Maximum**	**Average**	**Maximum**	**Average**	**Maximum**
Bofek_Contain	152.30	1107.65	0.81	12.56	0.83	12.06
Comb_Contain	231.53	1066.91	0.44	2.25	0.51	2.27
Spek_Contain	169.93	1066.91	0.51	6.01	0.45	3.11
Bofek_Intersect	8.46	1417.71	7.53	800.00	12.69	76148.66
Comb_Intersect	8.21	1270.03	6.45	46.30	10.07	17547.64
Spek_Intersect	8.28	1270.03	8.96	70.00	15.29	72470.59
Bofek_Within	3.92	246.69	9.10	698.80	29.21	76148.66
Comb_Within	3.50	151.53	10.20	46.30	27.65	17547.64
Spek_Within	5.07	179.46	12.26	64.71	35.05	72470.59

Table 15 - Taxation per selection method in Guldens

The previous tables show that there already some differences in size and taxation. For the last of this analysis, the land uses that are present in the selection will be evaluated and compared. This will be done for one methodology, in order to compare the different select by location options. The methodology of Spek and Pleijter is chosen to use in this analysis. This methodology is already proven in multiple other studies regarding raised bogs (F. De Vries et al., 2008; Grift, 2015; Worst, 2012)

The tables (16, 17, 18) below show the number of parcels that have a land-use and the average and total size of that land-use within the raised bogs areas. The Dutch names are used for the land-uses to make them correspond with the values in the cadastre. In the intersect method, the land-uses arable land (*bouwland*) and meadows (*weiland*) are the most common. At least, according to the number of parcels. When the total size is taken into account, these land-uses are only number three and four. Bogs (*veen*) and heathland (*heide*) are then much larger. Also, the average size of bogs and heathland are much larger than the meadows and arable lands. After the bogs and heathlands, the wetlands (*natte gronden*) have the largest average size. Other land-uses are less present in the selection methodology. Also, all land-uses are present in the intersect selection. The sizes in the *inhoudsgrootte* are larger than the size that calculate within the GIS. The relative difference between bogs and heathland and arable land and meadows is also smaller in the size with GIS attribute. This is related to the unknown land use class. This land-use accounts for fifteen per cent of all the lands within this selection method.

Spek_Intersect								
			inhoudsgrootte			**Size within GIS**		
land-use	**Number of parcels**	**%**	**Average**	**Total**	**%**	**Average**	**Total**	**%**
bouwland	10981	25.62	0.98	10801.65	5.61	1.92	21044.01	10.71
weiland	10395	24.25	1.43	14835.23	7.71	1.77	18438.03	9.38
heide	4639	10.82	19.72	91502.21	47.53	14.15	65647.1	33.41
veen	4242	9.9	14.89	63181.68	32.82	11.66	49452.47	25.16
hooiland	3613	8.43	1.44	5196.25	2.7	1.43	5161.34	2.63
water	3485	8.13	1.09	3785.43	1.97	0.98	3416.72	1.74
erf	1965	4.58	0.09	177.79	0.09	0.16	321.13	0.16
bos	1249	2.91	0.71	889.21	0.46	0.67	836.29	0.43
wegen	1001	2.34	0.58	577.5	0.3	0.77	771.93	0.39
natte gronden	554	1.29	2.56	1419.78	0.74	1.58	877.38	0.45
tuin	395	0.92	0.11	43.7	0.02	0.11	44.37	0.02
boomgaard	203	0.47	0.18	35.74	0.02	0.17	33.82	0.02
onbekend	77	0.18	0.83	63.93	0.03	395.13	30424.75	15.48
dijk	56	0.13	0.3	16.87	0.01	0.74	41.54	0.02
kerkelijk	7	0.02	0	0	0	0.34	2.39	0
Total (SUM)	42862	100	45	192527	100	432	196513	100

Table 16 - Land-uses and related size for the Spek_Intersect method

Spek_Contain								
			inhoudsgrootte			**Size within GIS**		
land-use	**Number of parcels**	**%**	**Average**	**Total**	**%**	**Average**	**Total**	**%**
Heide	66	55	508.56	33565	58.45	381.51	25179.71	30.89
veen	44	37	532.35	23423.47	40.79	492.59	21673.93	26.59
onbekend	4	3	0	0	0	7587.9	30351.6	37.24
weiland	3	3	146.29	438.87	0.76	1235.36	3706.08	4.55
bouwland	2	2	0.31	0.61	0.01	299.1	598.21	0.73
Total (SUM)	119	100	1188	57428	100	9996	81510	100

Table 17 - Land-uses and related size for the Spek_Contain method

Spek_Within								
			inhoudsgrootte			**Size within GIS**		
land-use	**Number of parcels**	**%**	**Average**	**Total**	**%**	**Average**	**Total**	**%**
bouwland	2668	28.36	0.71	1886.68	30.39	0.68	1817.82	38.57
weiland	2315	24.61	0.66	1530.16	24.65	0.6	1381.63	29.32
erf	1082	11.5	0.07	71.66	1.15	0.06	68.62	1.46
water	1045	11.11	0.48	506.52	8.16	0.12	127.62	2.71
heide	564	6	1.81	1021.16	16.45	0.95	535.94	11.37
hooiland	424	4.51	0.59	248.61	4	0.54	228.23	4.84
bos	399	4.24	0.33	130.11	2.1	0.27	107.97	2.29
veen	327	3.48	1.76	576.05	9.28	1.09	357.65	7.59
wegen	170	1.81	0.16	27.45	0.44	0.1	17.79	0.38
tuin	169	1.8	0.09	14.8	0.24	0.09	14.81	0.31
natte gronden	98	1.04	1.82	178.56	2.88	0.35	34.57	0.73
boomgaard	95	1.01	0.12	11.57	0.19	0.12	10.97	0.23
dijk	27	0.29	0.11	3.09	0.05	0.07	1.9	0.04
onbekend	21	0.22	0.1	2.14	0.03	0.32	6.76	0.14
kerkelijk	3	0.03	0	0	0	0.17	0.51	0.01
Total (SUM)	9407	100	9	6209	100	6	4713	100

Table 18 - Land-uses and related size for the Spek_Contain method

Also, all the land-uses are present. The most parcels are again arable land or a meadow. However, build-up areas (*erf*) and water are quite large as well. The bogs and heathland land-uses are less present in this selection method, at least in numbers of parcels that have this land-use. The average size of heathland and bogs are larger than the other land-uses, but the values are much lower than with the intersect selection. In total size, the heathland and bogs are smaller than the meadows and arable land in this selection, this is striking due to the fact that heathlands and bogs are known for their large areas (Gerding, 1995; Jong, 2009). This selection only selects 21 parcels that have an unknown land-use class. These lands are less than seven hectares in size and should not influence the results that much.

The last selection is the contain method. The land-use heathland and bogs are overwhelming in this selection. They account for 92 per cent of the parcels and for more than 99 per cent in size. Meadows and arable land are very small in this selection. However, the average size of land-uses is large compared with the other selections. The unknown class does not have any size value within the *inhoudsgrootte* attribute. However, when these unknown parcels are vectorised, they account for more than 37 per cent of the area.

To conclude, there are significant differences between the different selection methodologies. The BOFEK and Spek methodology do not cover the same area, with only a quarter that overlap. Not only the method influences the results, but also the select by location tool has a major influence on the findings, as shown with the land use analysis. Based on this first analysis, we could tell that the results differ too much to make concrete statements about raised bogs areas. Therefore, the areas will be analysed more in-depth in the following paragraphs. First, the raised bogs areas will be analysed per peat colony and the second case will focus on the property type. To compare results of these cases with the results of this paragraph, only the methodology of Spek and Pleijter is analysed in the remaining of the report.

5.2 Analysis of large and small bogs systems

This paragraph will focus on the third research question. This research question is: *How do small and large raised bogs differ in size, taxation, and land-use according to the cadastre of 1832?* As explained, the large bog systems are operationalised as the areas within the peat colonies. The smaller bogs are the areas that are not situated inside these areas. The peat colonies are obtained from the HISTLAND dataset. The HISTLAND dataset distinct five different peat colonies. These peat colonies will be compared to their attributes for one methodology, but three different select by location options. The methodology that will be used is the one of Spek (2004). The methodology of Spek is already proven in other studies regarding raised bog identification (F. De Vries et al., 2008; Grift, 2015; Worst, 2012). The within, intersect and contain options will be further analysed to also analyse the differences that occur due to the settings in the select by location tool.

5.2.1 Bourtanger Moor

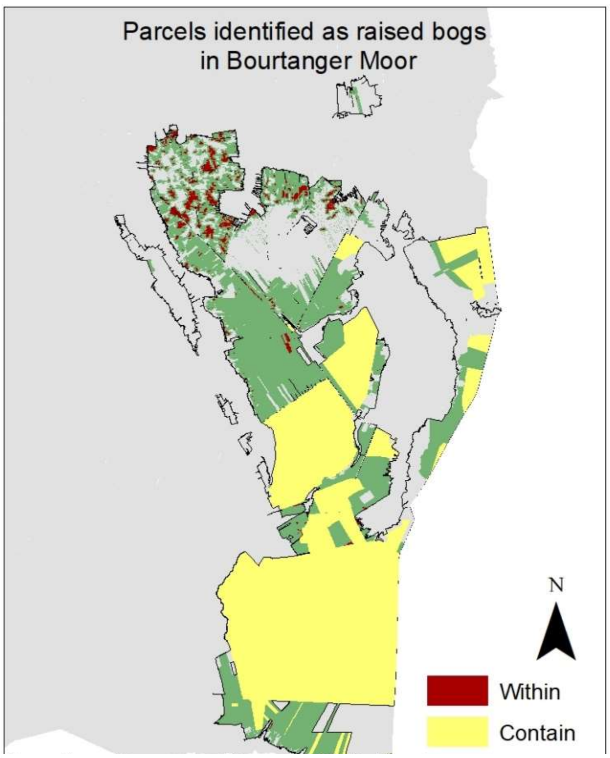


Figure 12 - Parcels identified as raised bogs in the Bourtanger Moor

The Bourtanger Moor was the most extensive system of raised bogs in the Netherlands (Casparie, 1993). Bourtanger Moor is divided into multiple shapes according to the HISTLAND dataset and is not one continuous area. In the HISTLAND dataset, the peat colony on the Bourtanger Moor is also the largest of the dataset. However, the parcels that are identified as raised bogs do not cover the complete area of the peat colony. Also, the number and spatial location of parcels that are assigned per methodology differs. The contain method only identifies twenty-seven parcels as raised bogs and are found in the southern part and around the border with Germany. The 2817 parcels of the within selection can be found in the northern part of the area. The intersect method identifies parcels in both the south and the north of the area but does not assign parcels in the west. The HISTLAND dataset identifies places as peat colonies, but according to the method of Spek and Pleijter, raised bogs are absent in that part.

Number of parcels		
Contain	**Intersect**	**within**
27	11487	2817

Table 19 - Number of parcels for the Bourtanger Moor

	inhoudsgrootte		**Size within GIS**	
	Average	**Total Size**	**Average**	**Total Size**
Contain	586.05	15823.29	1332.02	35964.47
Intersect	4.12	47353.79	5.53	63507.57
within	0.66	1857.06	0.59	1649.16

Table 20 - Sizes per methodology for the Bourtanger Moor

The size of the parcels differs between the multiple methods. However, the distribution of the values is the same as observed in the previous paragraph. The contain method has the largest average value for both the 1832 size as the size within the GIS. The average and total size of the area that is covered by raised bogs are significantly higher than initially calculated in 1832. In the intersect method, the size of the 1832 attribute is smaller than the GIS attribute. The difference, however, is less large than in the contain method. The average size only grows with one hectare and the total size with 15000 hectares, where the contain method grows more than 2 times its original size. The within size is the smallest of the three methodologies, in both average and total size. However, the sizes become even smaller, when the size within the GIS is calculated.

	Taxation per parcel		**Taxation per hectare (*inhoudsgrootte*)**	**Taxation per hectare (size within GIS)**
	average	total taxation	average per hectare	average per hectare
Contain	193.43	5222.6	0.34	0.42
Intersect	9.22	105929.7	10.44	14.4
Within	7.52	21194.52	14.04	28.99

Table 21 - Taxation within the Bourtanger Moor (in Guldens)

The terms of size, the differences between the three methodologies is large. In terms of taxation, the differences are less great between the Intersect and Within method. The average taxation per parcel is higher for the intersect method than for the within method. But the total taxation is larger for the parcels selected with the within method. The total amount of the contain method is the smallest of the three methodologies, but the average is also the highest. This is caused by the small number of parcels that lay within this selection. However, the average per hectare, which corrects for the number of parcels is the smallest for the contain method. Only 34 cents per hectare must be paid for these areas, where the other methods are above the ten Guldens. The values of the average per hectare become greater when the sizes of the vectorised parcels are used, but the distribution stays the same.

Table 22 shows the land-uses that are present in the selection. The size is calculated with the *inhoudsgrootte* attribute. This means that the parcels that have an empty or zero value as a size are neglected in this analysis. Therefore, all the unknown land-uses are neglected from the analysis. Dykes (*dijk*), religious grounds (*kerkelijk*) and unknown land-uses are not present in the raised bogs areas of the Bourtanger Moor. The contain method only have two land-uses in its selection: Bogs and heathland. Bogs account for more than 90 per cent of the area. The intersect method also identifies bogs and heathland as two prominent land-uses in the area. However, the percentages of the land-uses are lower. Also, arable land and meadows are often present in the selection. The within method do not account bogs and heathland as the most present land-uses in the Bourtanger Moor. The within method covers a smaller area, but slightly more than the half of the size is occupied by arable land.

	Intersect		Within		Contain	
Land-uses	**Total size**	**%**	**Total size**	**%**	**Total Size**	**%-**
boomgaard	13.72	0.03	2.9	0.16	-	-
bos	38.73	0.08	6.65	0.36	-	-
bouwland	4585.95	9.68	950.88	51.2	-	-
dijk	-	-	-	-	-	-
erf	45.11	0.1	23.95	1.29	-	-
heide	7499.49	15.84	239.2	12.88	1315.55	8.31
hooiland	444.19	0.94	13.66	0.74	-	-
kerkelijk	-	-	-	-	-	-
natte-gronden	30.53	0.06	1.39	0.07	-	-
onbekend	-	-	-	-	-	-
tuin	6.27	0.01	2.7	0.15	-	-
veen	31772.5	67.1	120.24	6.47	14507.74	91.69
water	287.78	0.61	77.47	4.17	-	-
wegen	141.64	0.3	1.99	0.11	-	-
weiland	2487.88	5.25	416.04	22.4	-	-

Table 22 - Land-uses and related size in the Bourtanger Moor

5.2.2 Hoogeveen

The area of Hoogeveen contains fewer parcels compared with the Bourtanger Moor. The number of parcels selected is around three times smaller than the Bourtanger Moor area. Again, the different methodologies are grouped over the area. The within method is located in the west of the area, were the contain lays in the north and east. The intersect combines the two methodologies. However, it should be noted that multiple raised bogs parcels surround parcels that are not a raised bog.

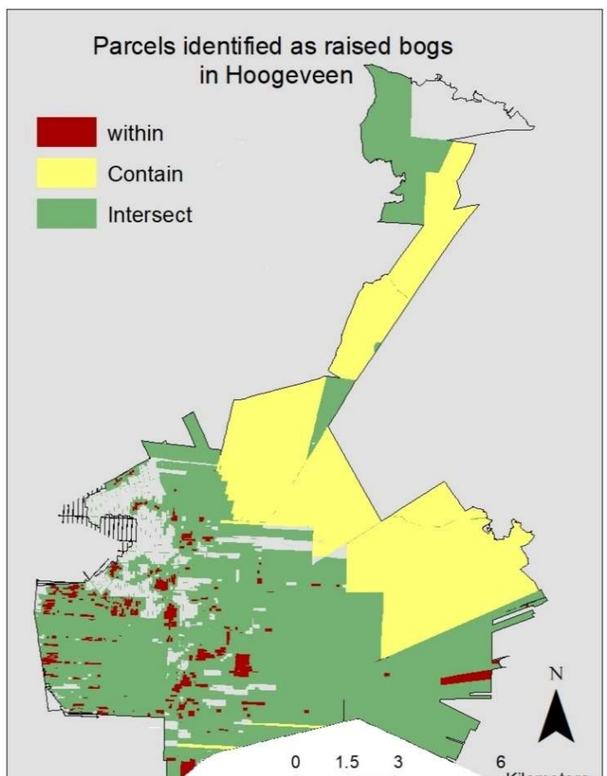


Figure 13 - Parcels identified as raised bogs in Hoogeveen

The average sizes in the *inhoudsgrootte* are comparable with the Bourtanger Moor area. The contain method is still the largest of the three. The contain method has an average size that is four hectares smaller than the counterparts of the Bourtanger Moor. The parcels in the intersect and within selection have an average size that is slightly larger than their counterparts in the Bourtanger Moor. The total size follows the same distribution as observed in Bourtanger Moor. The sizes become smaller when the size within GIS is used.

	inhoudsgrootte		Size within GIS	
	Average	**Total Size**	**Average**	**Total Size**
Contain	582.22	6404.39	357.39	3931.33
Intersect	5.25	18074.85	3.17	10913.8
within	0.75	694.36	0.38	350.02

Table 24 - Sizes for the three methodologies in Hoogeveen

Number of parcels		
Contain	**Intersect**	**within**
11	3442	932

Table 23 - Number of parcels per methodology in Hoogeveen

The taxation also follows the distribution as the Bourtanger Moor parcels. The contain have the highest average per parcel. This is significantly higher than in the parcels in the intersect and within methods. However, the average size per parcel is lower for the contain method than for the other methodologies. The size within the GIS makes the average higher for the other areas. This shows that the size of parcels is slightly smaller when they are measured within the GIS

	Taxation per parcel		**Taxation per hectare (*inhoudsgrootte*)**	**Taxation per hectare (size within GIS)**
	average	total size	average per hectare	average per hectare
Contain	251.04	2761.41	0.47	0.62
Intersect	3.37	11585.42	4.25	5.96
Within	1.53	1428.67	6.14	9.81

Table 25 - Taxation for Hoogeveen (in Guldens)

	Intersect		**Within**		**Contain**	
Land-uses	**Total size**	**%**	**Total size**	**%**	**Total Size**	**%-**
boomgaard	0.41	0,01	-	-	-	-
bos	218.98	1.21	62.97	9.07	-	-
bouwland	209.74	1.16	58.63	8.44	-	-
dijk	-	-	-	-	-	-
erf	8.03	0.04	3.18	0.46	-	-
heide	10190.13	56.38	180.72	26.03	4930.21	76.98
hooiland	4.07	0.02	-	-	-	-
kerkelijk	-	-	-	-	-	-
natte-gronden	-	-	-	-	-	-
onbekend	-	-	-	-	-	-
tuin	6.59	0.04	2.81	0.4	-	-
veen	7014.05	38.81	275.19	39.63	1474.18	23.02
water	140.61	0.78	25.21	3.63	-	-
wegen	29.41	0.16	7.05	1.01	-	-
weiland	252.85	1.4	78.6	11.32	-	-

Table 26 - Land-uses and related size for Hoogeveen

The contain method only contains two land-uses: bogs and heathland. However, the relative size is different compared to the Bourtanger Moor. The bogs are smaller (23%) were the heathland accounts for 77 per cent. In the within selection unknown, wetlands, religious grounds, hayland, dykes and orchards are not present. However, heathland and bogs are the largest land-use in this selection as well, but in this selection, the amount of bogs is larger than the amount of heathland. Other land-uses are less present compared with the Bourtanger Moor. For the intersect methodology, haylands and orchards are present, but the size of these land-uses is minimal. In the intersect method, the heathland account for more than fifty per cent of the total size.

5.2.3 Smildervenen

The third area that is investigated is the Smildervenen in the west of Drenthe. The Smildervenen are according to the number of parcels the third largest peat colony in the study area. The contain methodology only identifies two parcels as part of the Smildervenen area, which are found in the east of the area. The within and intersect methods identifies more parcels as raised bogs within the Smildervenen. The within parcels are located in the middle of the area near a central axis. The axis is the canal that divides this area. The intersect parcels are more present in the south of the area than in the north.

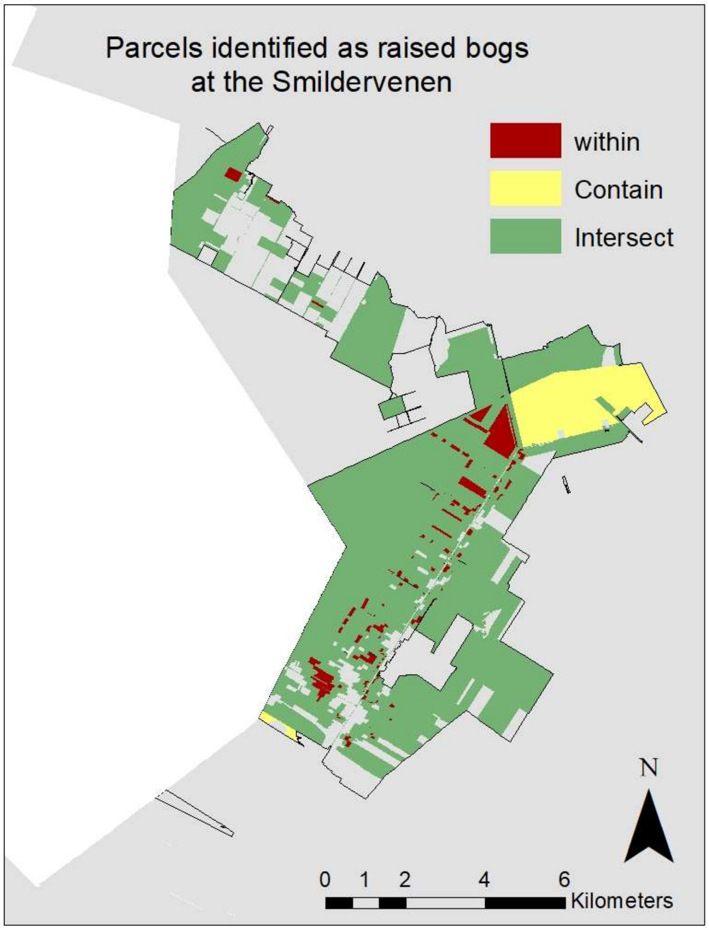


Figure 14 - Parcels identified as raised bogs at the Smildervenen

The two parcels that are selected in the contain method cover a relatively large amount of ground. This is comparable to the other areas that are observed. What is surprising is that the sizes between the attribute of 1832 and the recently calculated attribute do not differ very much. The total size of the area in the contain method is underestimated for only four hectares. These four hectares is in absolute numbers the highest differences. This is exceptionally low compared to the other areas that are already investigated. This could be explained by the fact, that there are no missing parcels in the area. Therefore, all the values in the attribute of 1832 are filled and do not necessarily influence the results.

The taxation values follow the distribution that is observed earlier in the report. The within has the smallest average per parcel, but the highest per hectare. This means that the parcels are relatively small but pay relatively much taxation compared with the other selections. The average taxation per hectare is again the lowest for the contain methodology. The average taxation per hectare differs between the attribute of 1832 and the attribute that contains the value of the GIS calculated size. The taxation becomes more for the intersect and within method.

Number of parcels		
Contain	**Intersect**	**within**
2	1813	362

Table 27 - Number of parcels at the Smildervenen

	inhoudsgrootte		Size within GIS	
	Average	Total Size	Average	Total Size
Contain	254.77	509.54	256.7	513.4
Intersect	3	5430.39	2.99	5427.86
within	0.6	216.68	0.59	215.18

Table 28 - Sizes per methodology at the Smildervenen

	Taxation per parcel		Taxation per hectare (*inhoudsgrootte*)	Taxation per hectare (size within GIS)
	average	total size	average per hectare	average per hectare
Contain	162.3	324.6	0.45	0.44
Intersect	4.38	7942.78	7.9	9.43
Within	2.52	910.44	12.14	19.82

Table 29 - Taxation per methodology at the Smildervenen (in Guldens)

	Intersect		Within		Contain	
Land-uses	Total size	%	Total size	%	Total Size	%
boomgaard	0.29	0.01	0.13	0.06	-	-
bos	9.14	0.17	1.58	0.73	-	-
bouwland	355.71	6.55	46.2	21.32	-	-
dijk	-	-	-	-	-	-
erf	6.85	0.13	3.44	1.59	-	-
heide	1126.44	20.74	29.78	13.75	16.48	3.23
hooiland	-	-	-	-	-	-
kerkelijk	-	-	-	-	-	-
natte-gronden	-	-	-	-	-	-
onbekend	-	-	-	-	-	-
tuin	13.08	0.24	4.97	2.29	-	-
veen	3578.32	65.89	100.28	46.28	493.06	96.77
water	97.9	1.8	0.04	0.02	-	-
wegen	61.24	1.13	0.51	0.23	-	-
weiland	181.41	3.34	29.75	13.73	-	-

Table 30 - Land-uses and related size at the Smildervenen

The table above shows the land-uses of the Smildervenen. The Smildervenen do not have the land-uses: haylands, dykes, religious grounds, wetlands and unknown. The latter is logical due to the usage of the *inhoudsgrootte* for calculating the size of the area. In all the methods, the bogs are the most present land-uses in the Smildervenen. Heathland and arable land are the two other prominent land-uses in the area.

5.2.4 Veenhuizen

The veenhuizen area is the smallest area in the HISTLAND dataset. It is located in the North-west of Drenthe close to the border of Groningen. Veenhuizen does not contain raised bogs areas according to the contain methodology. Therefore, the contain tables will not be present in the paragraph. The number of parcels for the within and intersect method are relatively close for this area. The intersect method have assigned 766 parcels as raised bogs against 357 parcels in the within method. The area is divided in two distinct parts. Most of the raised bogs parcels are found in the most eastern part.

Number of parcels		
Contain	**Intersect**	**within**
	766	357

Table 31 - Number of parcels per methodology at Veenhuizen

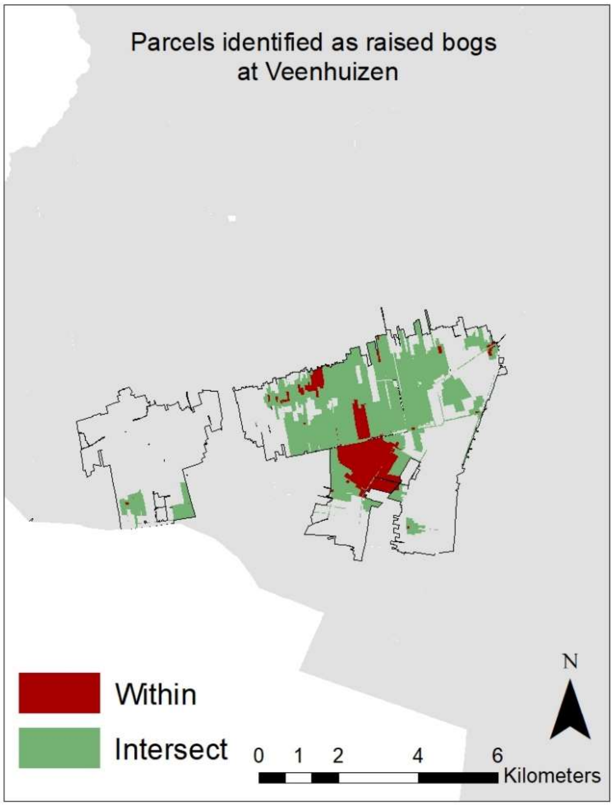


Figure 15 - Parcels identified as raised bogs at Veenhuizen

According to the total size, this area is the smallest of the five areas that are investigated. The sizes within a GIS are smaller than for the size that is measured in 1832. The average size of the intersect method is 1.84 hectares in 1832 and 1.62 measured within the GIS. The total area size is also lower than calculated in 1832. The same pattern can be observed for the area that is select with the within method. The within method is even smaller in total and average size. The average size is 0.73 in 1832, which is comparable with the size observed in Hoogeveen.

The taxation of these parcels is compared in table 32. The taxation per hectare is much larger for the size within GIS compared with the attribute of 1832. The taxation per hectare is almost two times higher when it is calculated with the size within GIS (17.4 Guldens versus 9.08 Guldens). The within method is also higher for the size within GIS. However, the difference is less than for the intersect method. The average per parcel is higher for the intersect method.

The land use table shows that the intersect and within method have the same land-uses in their selection. In both bogs are absent, but the heathland is mainly available. In both selections, this is the most present land-use. This could indicate that the area is not yet fully reclaimed and exploited. This is further stimulated due to the small amount of arable land and meadows in the area, because the heathland and peat are transformed into these two land-uses (Gerding, 1995; Van Beek et al., 2015).

	Taxation per parcel		**Taxation per hectare (*inhoudsgrootte*)**	**Taxation per hectare (size within GIS)**
	average	**total size**	**average per hectare**	**average per hectare**
Intersect	6.38	4889.4	9.08	17.4
Within	4.18	1493.8	13.31	17.73

Table 32 - Taxation per methodology at Veenhuizen (in Guldens)

	inhoudsgrootte		Size within GIS	
	Average	Total Size	Average	Total Size
Intersect	1.84	1408.36	1.62	1241.32
within	0.73	260.24	0.62	220.93

Table 33 - Sizes per methodology at Veenhuizen

	Intersect		Within	
Land-uses	Total size	%	Total size	%
boomgaard	1.09	0.08	1.09	0.42
bos	3.1	0.22	0.5	0.19
bouwland	149.8	10.64	41.49	15.94
dijk	-	-	-	-
erf	2.18	0.15	1.74	0.67
heide	1061.15	75.35	145.25	55.82
hooiland	-	-	-	-
kerkelijk	-	-	-	-
natte-gronden	-	-	-	-
onbekend	-	-	-	-
tuin	-	-	-	-
veen	-	-	-	-
water	71.77	5.1	18.57	7.14
wegen	22.19	1.58	7.26	2.79
weiland	97.08	6.89	44.33	17.04

Table 34 - Land-uses and related size at Veenhuizen

5.2.5 Odoornerveen

The peat colony near the Odoornerveen is located between the Bourtanger Moor and Hoogeveen. This area has the lowest number of parcels in the selection. With only three parcels in the contain method and six in the within methodology. The intersect contains 89 parcels, but this is still less than in the other areas. Despite the small number of parcels almost the full area is covered with raised bogs. Only some small parcels in the north and south are not covered by one of the methodologies. The contain selection only covers three parcels, but these parcels are the most prominent in the area. These parcels are quite large and located in the middle of the area. The within selection has some small parcels that are located in the middle of these contain regions. The additional parcels that are only selected with the intersect methodology are located near the border.

Number of parcels		
Contain	Intersect	within
3	89	6

Table 35 - Number of parcels per methodology in Odoornerveen

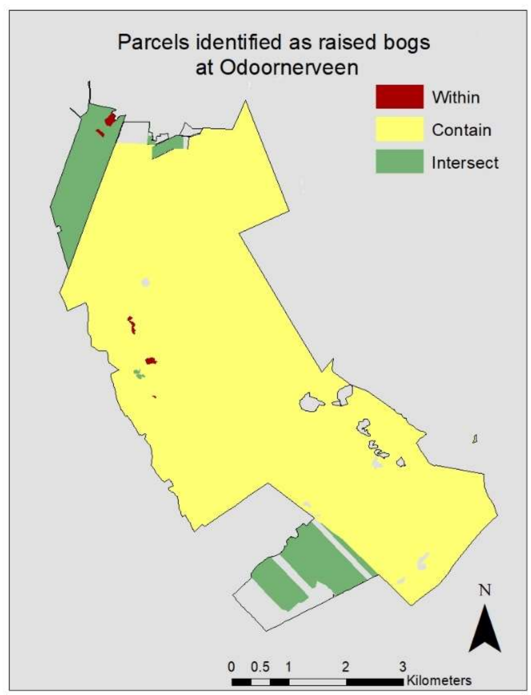


Figure 16 - Parcels identified as raised bogs at Odoornerveen

	inhoudsgrootte		Size within GIS	
	Average	Total Size	Average	Total Size
Contain	1086.96	3260.89	922.77	2768.32
Intersect	40.37	3593.1	34.92	3108.08
within	1.19	7.12	1.19	7.14

Table 36 - Sizes per methodology for Odoornerveen

The area differs much between these methods. The within parcels only cover seven hectares. The contain and the intersect method cover more than three thousand hectares. The average size also differs much between the three methods. The average size is only 1.19 hectares, where the contain method have an average of 1086 hectare. For this area, the size within GIS is smaller than the size calculated in 1832.

The taxation per parcel also differs significantly between the three methods. The contain method has an average taxation of 326 Guldens per hectare. This is the largest amount per hectare for all the areas that are investigated. The intersect is also the largest amount of all the areas. The taxation per hectare is low for every method compared with the other areas — however, the differences between the size in 1832 and the size in the GIS are almost non-existing.

	Taxation per parcels		Taxation per hectare (*inhoudsgrootte*)	Taxation per hectare (size within GIS)
	average	total size	average per hectare	average per hectare
Contain	326.09	978.27	0.3	0.4
Intersect	12.57	1119.02	0.56	0.55
Within	1.58	9.48	1.25	1.23

Table 37 - Taxation per methodology at Odoornerveen (in Guldens)

	Intersect		Within		Contain	
Land-uses	Total size	%	Total size	%	Total Size	%
boomgaard	-	-	-	-	-	-
bos	-	-	-	-	-	-
bouwland	-	-	-	-	-	-
dijk	-	-	-	-	-	-
erf	-	-	-	-	-	-
heide	1.68	0.05	-	-	-	-
hooiland	3.16	0.09	2.6	36.48	-	-
kerkelijk	-	-	-	-	-	-
natte-gronden	-	-	-	-	-	-
onbekend	-	-	-	-	-	-
tuin	-	-	-	-	-	-
veen	3565.69	99.24	1.01	14.24	3260.89	100
water	4.98	0.14	3.51	49.28	-	-
wegen	0.45	0.01	-	-	-	-
weiland	17.14	0.48	-	-	-	-

Table 38 - Land-uses and related size at Odoornerveen

A maximum of six land-uses is represent in this selection of the area of Odoornerveen. The broader intersect method contains all the classes. However, the bogs are represented for more than 99 per cent. The contain method only includes bogs. The bogs are much smaller in the within option. There they account only for 14.24 per cent. The land-uses hayland and water are larger in this selection.

5.2.6 The smaller bogs

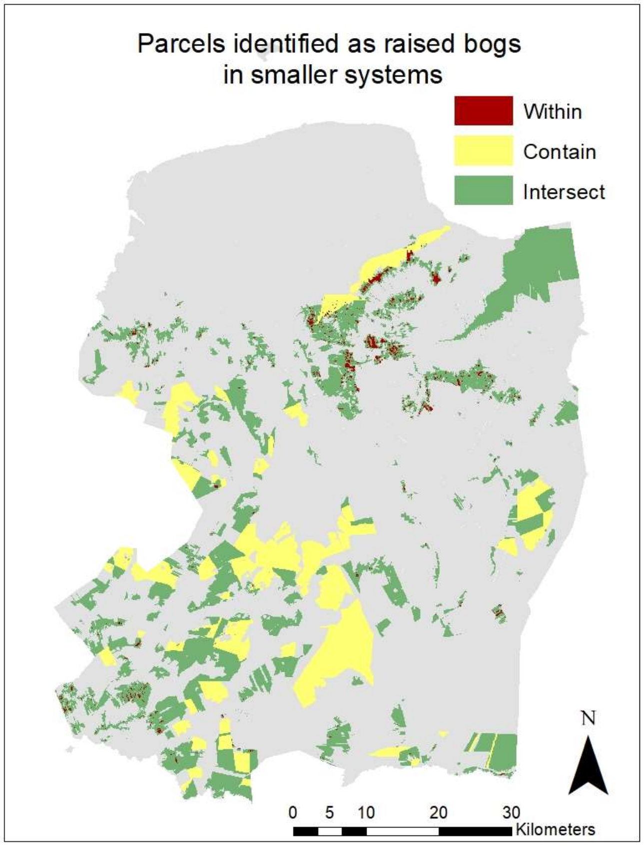


Figure 17 - Parcels identified as raised bogs in smaller raised bogs systems

This paragraph will investigate the bogs that are outside a peat colony. These raised bogs areas are identified as smaller bogs in the study area. The methodologies distinguish multiple smaller raised bogs. Various bogs lay nearby the peat colonies. Therefore, the question rises if the classification of the peat colonies can be this strict.

The number of parcels show that the smaller bogs contain more parcels than any peat colony. The within selection method identifies mainly parcels in the centre of Groningen and in the environment of Meppel. The contain method only identifies raised bogs areas in the centre of Drenthe and near existing peat colonies. The intersect method contains many more parcels, but the most striking is the parcels in the north-east. This is an area that lay outside the dykes of the mainland, but it is still classified as an raised bog area in this selection.

The size of the parcels follows the same pattern as in the peat colonies. The contain parcels are much greater than the within parcel, because, the intersect method contains both the contain and within parcels. the parcels the average size is slightly higher than the within methodology. The differences between the 1832 size and the size that is recently calculated is quite small for the intersect and within method. The contain method has some differences. The average size is almost hundred hectares higher for the GIS size. The total size is also higher, this can be caused do to the presence of missing parcels. These missing parcels are the parcels that are not digitalised due to the absence of the OAT.

	inhoudsgrootte		**Size within GIS**	
	Average	**Total Size**	**Average**	**Total Size**
Contain	413.55	31429.84	504.37	38332.01
Intersect	4.62	116666.5	4.45	112314.6
within	0.64	3173.1	0.46	2270.36

Table 40 - Sizes per methodology for the smaller bogs

Number of parcels		
Contain	**Intersect**	**within**
76	25265	4933

Table 39 - Number of parcels per methodology in the smaller bogs

	Taxation per parcel		**Taxation per hectare (*inhoudsgrootte*)**	**Taxation per hectare (size within GIS)**
	average	**Total taxation**	**average per hectare**	**average per hectare**
Contain	143.88	10934.66	0.58	0.43
Intersect	8.85	223601.7	9.05	17.37
Within	4.6	22703.04	12.36	45.7

Table 41 - Taxation per methodology for the smaller bogs (in Guldens)

The taxation has also the same distribution as the peat colonies. The contain method has the highest taxation per parcel, but the lowest taxation per hectare. The within has the lowest taxation per parcel, but the highest per hectare. The size within GIS reduces the amount of taxation paid per hectare for the contain method.

All land-uses, except religious grounds, are present in the methodologies. Heathland is by far the most common land-use in the intersect method followed by bogs and meadows. In the within selection, arable land and meadows are the largest land-uses. However, bogs and heathland are still of a considerable size. The contain method only hold heathland or bogs, with only a small portion of meadows. In both the intersect and within are parcels present that do have a size in the 1832 attribute, but that have an unknown land-use. The missing parcels (thus without an OAT) are not represented in table 42, because the missing parcels have both an unknown land-use and an unknown size. This unknown size results in an empty or zero. Which means that the per centage will be zero and thus non-existed.

	Intersect		**Within**		**Contain**	
Land-uses	**Total size**	**%**	**Total size**	**%**	**Total Size**	**%**
boomgaard	20.24	0.02	7.46	0.24	-	-
bos	619.26	0.53	58.41	1.84	-	-
bouwland	5500.46	4.71	789.47	24.88	-	-
dijk	16.87	0.01	3.09	0.1	-	-
erf	115.62	0.1	39.34	1.24	-	-
heide	71623.31	61.39	426.2	13.43	27302.76	86.87
hooiland	4744.83	4.07	232.36	7.32	-	-
kerkelijk	-	-	-	-	-	-
natte-gronden	1389.25	1.19	177.17	5.58	-	-
onbekend	63.93	0.05	2.14	0.07	-	-
tuin	17.76	0.02	4.32	0.14	-	-
veen	17251.12	14.79	79.34	2.5	3687.6	11.73
water	3182.4	2.73	381.72	12.03	-	-
wegen	322.56	0.28	10.65	0.34	-	-
weiland	11798.87	10.11	961.43	30.3	438.87	1.4

Table 42 - Land-uses and related size for the smaller bogs

6. Discussion

Current literature does not provide detailed information about the quality assessment of this kind of, historical geographical, datasets. Also, potential users of historical, geographical datasets lack the skills and knowledge to do a proper quality assessment. Users of the historical, geographical datasets skip the quality assessment of their data and start at once with their analysis. This study aims to raise awareness of the importance of quality assessment in historical GIS. This is done by offering methodologies that can be applied on historical, geographical datasets. The GIS version of the Dutch cadastre of 1832, a popular source in historical science, is used as a subject for the quality assessment. After the quality assessment, the cadastre is used to study raised bogs for a larger study area. The second and third objective had a focus on the study of these raised bogs areas. The second objective tries to develop a methodology for raised bogs identification. As shown, multiple methods can be used for the identification. The last part of this study had a specific focus of raised bogs areas in various locations. Do larger raised bogs systems have different cadastral attributes than smaller bogs systems?

The first part of this chapter will focus on the interpretation of the results. How could the values be explained and how does it relate to existing literature? Afterwards, the limitations of this research are discussed, and some recommendations for further research are provided.

6.1 Interpretation of the results

The interpretation of the results will be provided per part of the research. This means that first the quality assessment results will be discussed. This will be followed by the results of the comparison of the identification and selection methodologies. The last part of this paragraph will focus on the comparison of the raised bogs areas.

6.1.1 Quality assessment

The quality assessment was done for the cadastral dataset of 1832 of the provinces of Drenthe and Groningen in the north of the Netherlands. The literature review identifies multiple elements that indicate the quality of a geographical dataset. However, some elements are more critical in historical GIS (the application of historical datasets in GIS). Users of historical GIS focus on the attributes and completeness of the data (Gregory & Healey, 2007; Knowles, 2005). The quality assessment offers simple methodologies for the assessment of these elements. The methodologies need only basic knowledge of GIS. This level of knowledge is often available by users of historical GIS (Knowles, 2005, 2014). Therefore, these methodologies could also be applied to other historical, geographical datasets.

The results of the quality assessment also provided some interesting findings. The accuracy of the attributes was considered well according to the error matrix (table 7). Only 1.53% of the features contained an error, which was most of the time an incorrect interpretation of the handwriting in the original registers. Other elements like the completeness are less well. Some of the municipalities in Drenthe did not have any attributes filled, because of the absence of the OATs. However, there is observed that some of these OATs are available on the *Beeldbank* (Rijksdienst voor het Cultureel Erfgoed, n.d.). A reason could be that the scans of the OATs became available after the completion of the project of Drenthe because the project shows that these OATs were absent.

The positional accuracy was measured with a comparison of the size as calculated in 1832 and the size of the vectorised parcel. Ideally, the differences are minimal between these two attributes. However, it is seen that especially the networks (roads, waterways, and dykes) are influenced by more significant differences, where the size in 1832 is much smaller than the size that is calculated in the HISGIS project. The difference could occur due to the different institutions that have taken part in the digitalisation of the provinces. Based on this information, it was expected that the total size of the digital cadastre was larger than the size in 1832. However, the opposite was the case. The study area was more than 800 square kilometres larger in 1832 then is calculated in this study. The reason for this does not appear in this study. However, it seems unrealistic that the calculation methods in 1832 were this inaccurate.

The usability appeared to be well. One attribute had to be standardized, because of the difference in units that were used. Also, the merging of the datasets was an easy task. However, it was harder to link the attributes of the two datasets. This means that is hard to know what the attribute stored in the dataset of Drenthe and which one was his counterpart in the Groningen. The names were not for all attributes the same. Also, a brief explanation in the metadata could be useful, especially for non-Dutch speakers.

In general, the quality of the dataset is high and can provide interesting results when it is applicated to a specific use-case. The attributes could contain some strange values. Land-uses are misspelled, sizes are not always what they should be according to the positional accuracy test. Therefore, the users of the data should always try to motivate, how the results are generated. They should not accept the outcome immediately as the truth.

6.1.2 Impact of identification- and selection methodologies

The second part of the study specifically focussed on the application of the cadastre on raised bogs areas. The identification of raised bogs in 1832 are a vital part of the analysis, because of the absence of a dataset containing raised bogs in 1832, which should be constructed. The distinction of these raised bogs soils was performed with three different methodologies. These methodologies indicate which areas are raised bogs and which areas are not. The next step was to assign parcels to these methods with the select by location tool. The assignment is done with three different options (contain, intersect and within). Based on the results, it can be argued that the selection method and selection option have a significant impact on the final results with different values for the taxation, size and land-uses in the selections.

The BOFEK and SPEK methodology do not cover the same areas. The analysis shows that less than 23% of the parcels overlap (the combined methodology). The methodology of Spek (2004) was already proven in scientific literature (Grift, 2015; Worst, 2012). However, the methodology does not provide the bogs in 1832, but the areas that were bogs in history. However, it enables the opportunity to observe how these bogs were used before, during and after the reclamation and exploitation of the area. The BOFEK methodology is constructed with the help of literature relating raised bogs. However, the BOFEK map is created on the basis of the soil that is currently visible (Alterra, 2018). Bog soils that were present in 1832 could have vanished over the years through oxidation or exploitation. This means that the bogs soil is not visible in the soil of nowadays.

Also, the select by location tool has a significant impact on the results. Where the contain option has a large average size and a low taxation per hectare. The within option has the opposite, a relatively small size, and a high taxation per hectare. The intersect method has values that lay between these two. This can be explained, that the intersect method includes both options, plus additional parcels. The values in size can be explained due to the nature of the select by location tool. The within option means that a parcel should be within a raised bog area. This results in small parcels that are identified as raised bogs. The contain method works the other way around; the parcels should cover the raised bog area. This results in relatively large parcels, because they cover more easily the raised bogs areas.

The comparison of the land-uses shows that the different methods provide different average sizes. The general trend that occurs is that heathland and bogs were relatively large areas. The heathland parcels were on average even larger than the bogs. Multiple reasons can explain this. First, the content of the bog category is quite diverse. This category includes exploited parcels, which are already divided and almost ready for a transformation to another land use, like arable land or a meadow. These small areas lower the average size of the bog category. In a dynamic area, like the province of Drenthe, the number of this kind of parcels can be quite high. Another reason is that these heathlands were a vital part of the agricultural society. The dung of animals that lived on these areas was used to make the exploited peatlands more fertile. This could mean that the heathlands were common ground and that these fields were not divided into smaller plots of land with multiple owners. This makes it easier to ensure to keep the regulations of these areas. Heathland was not transformed to another land-use to ensure that the exploited parts became more fertile (Hartog, 1884). Hartog told already that the heathland was later reclaimed than the other surrounding areas and this is further supported by Thissen (n.d.). According to Thissen, more than 33% of Drenthe was still heathland in 1833. This statement cannot be thoroughly checked with this study, because not all raised bogs were heathland and vice versa. However, the results show that the heathland account for more than 33% of the land-use in our study.

The comparison of the selection methodologies provides multiple views on the study area. The contain methodology shows results that are most in line with the view of common literature. The raised bogs areas are not reclaimed and exploited. It is resulting in large parcels that are still bogs or used as heathland with a low taxation per hectare. This could be an indicator that there was no competition for these lands. However, the results could also provide an entirely different view of the study area. The within method shows that the study area is already quite advanced in the exploitation of the bogs. More than 50% is exploited and transformed into arable land and meadows. There are still some heathland and bogs in Drenthe. However, these areas are quite small. Literature suggests that both views are inaccurate. Gerding (1995) describes that large-scale exploitation in Groningen and Drenthe had started in the 16th century. The large exploitations lasted until the end of the 19th century. Therefore, it seems that the within and contain method provides inaccurate views of the study area.

The intersect methodology of Spek provides a more accurate view on the study area. This method shows that some parts of the raised bogs are already reclaimed and exploited. This can be observed due to the presence of arable land and meadows. However, this methodology also shows a large amount of bogs and heathland, which still needs to be reclaimed and exploited. The large amount of the land-use bogs can indicate that the actual exploitation sites are covered in this land-use. Literature suggests that the large-scale exploitation of peat areas was on the highest point in the 1800s (Gerding, 1995; Knol & Noordman, 2003; Van Beek et al., 2015).

The large part of the analysis is only performed with the methodology of Spek, because this literature was already proven in other scientific studies (F. De Vries et al., 2008; Grift, 2015; Worst, 2012). Therefore, it is expected that this methodology provides the most accurate results for the raised bogs identification compared with the BOFEK and combination methodology. The options of the select by location tool offers also different results. It was expected that already a part of raised bogs in the study area was exploited. However, also at least 30% of the study area should be heathland (Thissen, n.d.). This view is most closely met with the intersect methodology. Therefore, it can be argued that the methodology of Spek with the intersect option offers the best results of the nine methodologies that are compared in this research.

6.1.3 Comparison of large and small raised bogs systems

There was chosen to distinguish large and small bogs systems, because information about the latter less studied. The cadastral dataset offers the opportunity to compare various large and small areas in one large analysis with the same methodology. The areas are still compared with three methodologies to see if patterns occur in all selection methods. The three methodologies serve as a validation of each other. It prevents that a pattern occurs that is a result of a subjective choice for a selection methodology. When a pattern occurs in multiple methods, there can be argued that the pattern actually exists.

The results show that there are some significant differences between the areas. Notably, the area of Veenhuizen and Odoornerveen do not fit in the pattern of the others. Veenhuizen does have large spaces that have the land-use heathland. However, there are no parcels with bogs in the area. This could mean that the area was not in the process of exploitation or reclamation. The thesis of Gerding (1995, pp. 212–216) shows a detailed description of the bogs near Veenhuizen. The peat production in this area was minimal in the year of 1834-1864. However, there was still some production, and these production sites are not visible in the cadastral datasets. An explanation of the high amount of heathland is also provided. The process of reclamation was slow in the area and lasted until the second World War. There are also still some raised bogs present in the area that are never reclaimed. It indicates that a large part of Veenhuizen was not reclaimed in 1832 and that this could be used as heathland.

The Odoornerveen stands out for multiple reasons. The size of the parcels is relatively large in the contain and intersect methodology. The taxation per hectare is also extremely low, and the dominant land-use is the bog category. All of this suggests that the area is not reclaimed and exploited in 1832. A literature review on the Odoornerveen shows that this area was indeed not in the process of exploitation (Gerding, 1995, pp. 218–224). The main canal, the central axis of the area, was created in 1853 and it was not until 1855 that the peat production started. It is also remarkable, that the thesis of Gerding identifies Veenhuizen and Odoornerveen as smaller bogs instead of large areas.

The other areas have fewer results that stand out. The smaller bogs have relatively much heathland compared to bogs and other land-uses. This could show that the smaller bogs are also not yet in the state of exploitations. However, it is hard to find information that confirms or contradicts this statement. However, it is surprising that the taxation per hectare is on the same level as the other three areas (Smildervenen, Hoogeveen and the Bourtanger Moor).

These other areas have comparable results in the land-use distribution of bogs and heathland. The Bourtanger Moor has a larger space that is classified as arable land compared to Smildervenen and Hoogeveen. This could also explain why the taxation is slightly higher per hectare in the Bourtanger Moor than the other areas.

6.2 Limitations

The study also had some limitations. The most important limitation is the suitability of the Dutch cadastre of 1832 for raised bogs analysis. As discussed in parts regarding the quality is that the quality is quite well, but the completeness of the data and the usability lower the suitability of this dataset. Five municipalities are not (entirely) present in the digital representation of the cadastre. The location of these missing municipalities is nearby or on raised bogs soils (Casparie, 1993; F. De Vries et al., 2008; Gerding, 1995; Pleijter, 2004; Spek, 2004). This means that the results are biased, because not all the parcels of 1832 are analysed. However, the absence of data could also be related due to the presence of raised bogs. These areas were hard to cross, and this does not help when an area is mapped by a surveyor (Baaijens, 2002).

Also, the attributes of the cadastre are not entirely trustworthy. Many attributes have nominal values, like names, professions, and place of residence. This makes a proper analysis hard, and therefore a reclassification should be applied to group similar values. However, reclassification can be time-consuming and subjective. In this research, there is profited from reclassification work of HISGIS, which resulted in little manual labour. The automatic classification increases the necessary GIS skills, which are not always present by the historical GIS users (Knowles, 2014). Also, the ratio attributes like size and taxation are not entirely trustworthy. The size of many parcels differs much between the size calculated in 1832 and the size that is automatically calculated in GIS. The valuation and taxation of areas differ per cadastral municipality (Veldhorst, 1991). This means that farmland of medium quality is not evaluated the same. Farmland of medium quality in an area with many low-quality fields will be valued much higher than a medium quality field in an area with high-quality fields. The problem is tackled with the assumption that raised bogs areas are always relatively low in valuation. The actual taxation value is not used for conclusions, but more the general trend that occurs through the taxation values.

The research aimed to clarify the impact of a selection method and not to select the best methodology. The methodology of Spek is already proven in literature (Grift, 2015; Worst, 2012) and could be therefore the best choice. However, the impact of the select by location tool, that is used to combine the dataset of raised bogs with the cadastre of 1832, is also significant. Also, for this, the best choice could not be made based on this research. However, the research shows that the intersect method in combination with the Spek methodology offers the most representative results in this study. But, the intersect option could be too broad and identifies parcels too easily as raised bogs. A slightly stricter method is the 'centroid' method. The centroid method assigns a parcel as a raised bog when the centre of a parcel lay inside the polygon of the raised bog area. This option is not studied in this research but could help in further research regarding raised bogs and GIS.

The sample size of the attribute accuracy held only 100 parcels per dataset. This amount is chosen for practical reasons, and with this amount of hundred parcels it still reaches the 99% confidence level, however, the confidence interval is quite high with 10%. A larger sample size provides a more detailed view of the actual attribute accuracy. Now the results could be biased.

This study has tried to assess a historical, geographical dataset. However, the tools available for quality assessment are not all suitable for this type of data. The assessment of modern source includes the usage of reference data. But, with historical data, there is often no reference available (Gregory & Healey, 2007; Knowles, 2002). This study has created some methodologies that could substitute traditional quality assessment methods in case historical data is assessed. The methodologies should also not require too much GIS skills in order to understand by the historical GIS users easily. These users can then reproduce the procedures on their dataset and provide a quality assessment and how it influences their results. However, the methodologies were not all science-based.

The third research question identifies smaller bogs as bogs outside the peat colonies. This is not entirely true, because many peat colonies exist. Almost all the villages that were founded in the raised bogs areas are peat colonies. The smaller bogs group also contain some larger raised bogs like Bargerveen and Fochteloërveen. These areas are quite large, and never exploited and reclaimed. However, it is therefore unrealistic that these areas are called small raised bogs systems. Also, the research shows that there many smaller bogs in the study area. They were analysed as a group. However, it is not clear if they share the same characteristics. Some smaller bogs should be earlier exploited than others. However, how many of the smaller bogs are exploited cannot be distinguished from this research.

6.3 Application of the methodology in other studies

This paragraph will highlight if the methodology of this research can be applied to other areas, timeframes, and cases. Multiple arguments can be given why the quality assessment methodology in principle be applied to other historical geographical datasets. Firstly, the attribute accuracy uses an error matrix to check the errors from datasets, which are digitalised form an analogue source. Secondly, the completeness and usability check, which uses an count operation, can be applied easily on other datasets. At last, the topology checker can be applied to other historical datasets, because the rules that can be checked are not limited to gaps and overlaps of features.

However, the positional accuracy cannot always be check in the same way as is done in this research. Two different size attributes are not always present in other datasets. However, some dataset could have alternatives that have similar data. These datasets can then be used a reference for more advanced positional accuracy checks. For example, if the location of the features is on the right spot. This is especially true for features that are visible in the current landscape, like churches and buildings.

The methodology of the analysis for the raised bogs areas is not case specific. The analysis of the cadastre is a summary statistics operation, that can be applied on all datasets with attributes. It does not matter if the features are polygons, lines, or points. When the cadastre of 1832 is used, various areas and cases could be used. For example, the cadastre of 1832 could be analysed in the province of Limburg or in the city of Amsterdam. Another examples could be the investigation of religious grounds (*kerkelijke gronden)* or the disappearance of forests in the Netherlands. The information is stored in the cadastre and with help of the same operations the information can be analysed. When the cadastre of 1832 is used, the methodology cannot be applied for other time periods. Also, for cadastral data of other years, the methodology could be applied to study raised bogs.

7. Conclusions and recommendations

7.1 Conclusions of research questions

As stated at the beginning of this report, the main objectives were to assess the quality of the digital version of the Dutch cadastre of 1832 and to applicate the cadastre in research regarding raised bogs. To achieve these objectives, four research questions were formulated, which will be answered in this chapter.

1. *What metrics determine the quality of geographical data in Historical GIS and how can these metrics be operationalised to assess the quality of the historical GIS data?*

The metrics that determine the quality of geographical data can be grouped into two groups of quality. The first group is the internal quality, which is related to the characteristics of the data itself. In order to measure the internal quality, the following seven metrics are distinguished by Van Oort (2006). The internal quality metrics are Lineage, Positional accuracy, Temporal accuracy, Attribute accuracy, Completeness and Usability. The external quality focuses more on the fitness-for-use and has three different metrics: Discovery and re-use of data, Data format and the use-case. The main problem with these metrics is that some do not apply to historical, geographical data or that they are hard to measure. Some metrics need reference data for comparison, and these are not available. Therefore, the operationalisation skips some of the metrics. Also, metrics like completeness and attribute accuracy are more critical when using historical data.

The primary focus in the operationalisation was that it was reproducible by historical GIS users. These users know how to interpret the data, but do not have many GIS skills (Knowles, 2014). The methodologies of quality assessment do not contain advanced methodologies, but checklists, count and math operations.

2. *What is the quality of the GIS version of the Dutch cadastre of 1832?*

The quality of the GIS version of the cadastre of 1832 was assessed with the help of six of ten metrics of quality assessment. The study area contained two datasets, one for Groningen and one for Drenthe. The most important metrics were completeness and attribute accuracy were assessed first. Despite the small sample size, the accuracy of the attributes was considered good, where the Groningen dataset was slightly better than the Drenthe dataset. In terms of completeness, the Groningen dataset was complete, but the dataset of Drenthe lacks several municipalities. Unfortunately, these missing were also found on spaces with potentially raised bogs. The logical accuracy, which checked for gaps and overlaps between parcels, found many topology errors. The positional accuracy compared the size calculated in 1832 with the size that is calculated within the GIS. During this analysis, many parcels were found with a difference of more than 17%. Most of these parcels were related to roads, waterways, and dykes. The evaluation of the metadata (internal usability) showed that all necessary elements (Geonovum, 2016) were present. The external usability had a focus on the standardisation, merging and reclassification of the dataset. The first two operations were simple and only slightly hindered by differences in the names of the attributes. The reclassification was already done for Drenthe. The same reclassification scheme is applied to the dataset of Groningen.

To summarize, the quality of the dataset of Drenthe is less than the quality of the dataset of Groningen. The internal usability was the same for both datasets as well as the distributions of errors in the positional accuracy test. The logical accuracy and the external usability were better for the Drenthe dataset compared to the dataset of Groningen. However, Drenthe scores less on the two most vital metrics for historical, geographical data. Therefore, the quality of the dataset of Drenthe is assessed less well than the Groningen dataset. In general, the quality of the dataset of the study area is high.

3. *How can raised bogs areas in 1832 be distinguished in a GIS and what is the impact of different methodologies?*

There is no dataset available that holds the raised bogs areas in 1832. However, multiple datasets do contain information about the soil, and these could be used as a substitute. In this research, two methodologies and a combined version are compared. The methodology of Spek is already proven by other studies (Grift, 2015; Worst, 2012). The BOFEK methodology is constructed with the help of existing literature. The combined version contains the areas are both identified as raised bogs in the two other methodologies. These methodologies are assigned to the cadastral dataset with the select by location tool. This tool has multiple options to select parcels that are in a raised bog area.

The research shows that mainly the select by location tool has a significant impact on the results. The taxation, sizes and land-uses in the area differ much between the three options that are investigated. The differences between the selection methodologies also differ much. They both identify different areas as raised bogs areas. The research did not focus on the best selection method, but because the method of Spek is already proven in scientific studies, this one appears to be better in identifying raised bogs areas in the past.

4. *How do small and large raised bogs systems differ in size, taxation, and land-use according to the cadastre of 1832?*

The large raised bogs were identified with the help of the HISTLAND dataset, which contained five larger peat colonies. The small bogs were identified as areas that lay outside these areas. The results show that the smaller bogs do not stand out compared three of the five-colonies. The smaller bogs have similar average sizes, the taxation per hectare is on the same level, and the same land-uses are present in the same distribution as the Bourtanger Moor, Hoogeveen and the Smildervenen.

Odoornerveen and Veenhuizen stand out in terms of the land-use. Veenhuizen was already a long time a peat colony, which is shown in the cadastre through the presence of build-up, meadows, and arable land. Bogs are not present in Veenhuizen. However, other research suggests that peat production sites should be the presence (Gerding, 1995, p. 216). The Odoornerveen stands out, because of all the attributes. The dominant land-use is bogs, the taxation per hectare is low, and the average sizes of parcels are large. The Odoornerveen is not reclaimed and exploited in 1832 (Gerding, 1995, pp. 218–224). This shows that the attributes of the cadastre can tell when an area is exploited.

7.2 Recommendations for further research

The following recommendations for further research are given:

- Use Natura2000 areas for raised bogs identification

The Natura2000 areas contain the raised bogs areas that are still present in the landscape. These areas were also raised bogs in 1832 and this drops the uncertainty of the Spek and BOFEK methodology, if the identified areas where truly raised bogs. The research can also provide more insight how the remaining raised bogs areas were used in 1832 and if this differs from other areas that are studied in this research

- Research of raised bogs with the Spek methodology and centroid option

Historical research on raised bogs areas with the help of the cadastre of 1832. When immediately a choice is made regarding the selection method and the option in the select by location tool more in-depth research can be conducted. This provides more results about the raised bogs areas in 1832. This could help to verify the results of this research. Could also be done for a specific case like a small raised bogs system.

- Analyse the land ownership of raised bogs areas

This study neglects the ownership of a parcels. However, the cadastre offers four attributes related to the owner of a parcel: surname, first name, profession, and place of residence. These attributes can provide more insight who owns a parcel and there can be observed if raised bogs areas are owned by a group of similar people. However, during this study it is observed that these attributes should be heavily processed due to inadequate filling of the cadastre in 1832. Sometimes, names of companies are split between first and surnames. Also, the attributes should be reclassed to analyse the results more efficient in a GIS.

- Methodologies for quality assessment of historical, geographical datasets

Research into different methodologies of quality assessment of historical, geographical dataset. The methodologies should be understandable and usable for the occasional GIS users, because scholars of historical sciences do not have many GIS expertise (Knowles, 2014). This could increase the quality of these studies and makes the findings more trustworthy.

- More usage of the digital cadastre of 1832

Despite the limitations of the cadastre as highlighted in the limitations and in the article of Veldhorst (1991), the cadastre can be a good source for further research in 1832. Knol and Noordman (2003) combines the cadastre with the height model of the Netherlands. However, more historical research could profit from the contents of the cadastre.

References

Ablebits. (n.d.). Ultimate Suite for Excel. Retrieved January 9, 2019, from https://www.ablebits.com/excel-suite/index-2018.php

Agumya, A., & Hunter, G. J. (1999). A Risk-Based Approach to Assessing the "Fitness for Use" of Spatial Data. *URISA Journal*, *11*(1), 33–44. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.2953&rep=rep1&type=pdf

Alterra. (2006). Bodemkaart van Nederland, schaal 1:50.000, met veenkartering, Versie 2006. Retrieved December 4, 2018, from https://nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/1032583a-1d76-4cf8-a301-3670078185ac

Alterra. (2018). Bodemfysische Eenhedenkaart (BOFEK2012) - WUR. Retrieved December 5, 2018, from https://www.wur.nl/nl/show/Bodemfysische-Eenhedenkaart-BOFEK2012.htm

Baaijens, G. J. (2002). Wolven als aanwijzers voor verdroging in Drents hoogveen. *De Levende Natuur*, *103*(5), 164–164. Retrieved from http://natuurtijdschriften.nl/download?type=document;docid=495373

Berendsen, H. J. A. (2008). *De vorming van het land - Inleiding in de geologie en de geomorfologie*. *Fysische geografie van nederland*. Koninklijke Van Gorcum. Retrieved from https://books.google.nl/books?hl=nl&lr=&id=nkUhG13HX1kC&oi=fnd&pg=PA1&dq=berendsen+vorming+van+het+land&ots=UsTYTxBL4y&sig=zaWWNO7gKL1oc7lbs2O91CJvaRE#v=onepage&q=berendsen vorming van het land&f=false

Bont, C. de. (1997). *Gegist bestek: Proefonderzoek van de Gelderse Vallei en omgeving voor een Historisch-Geografisch GIS*. Wageningen. Retrieved from https://library.wur.nl/WebQuery/wurpubs/fulltext/299495

Bont, C. de. (2004). The Significance of the Dutch Historical Gis Histland. In *European Rural Landscapes: Persistence and Change in a Globalising Environment* (pp. 345–358). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-0-306-48512-1_21

Bont, C. de. (2009). Middeleeuws Veen en Middeleeuwse Veenboeren. *Grondboor & Hamer*, (3/4), 85–93. Retrieved from http://natuurtijdschriften.nl/download?type=document;docid=568316

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., & Rampini, A. (2014). A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. *Information Sciences*, *258*, 312–327. https://doi.org/10.1016/j.ins.2013.07.013

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., & Rampini, A. (2016). On predicting and improving the quality of Volunteer Geographic Information projects. *International Journal of Digital Earth*, *9*(2), 134–155. https://doi.org/10.1080/17538947.2014.976774

Bos, J. (2017). *Locatie en bezit van Hoofdelingen en Eigenerfden in Laatmiddeleeuws Friesland*. Universiteit Leiden. Retrieved from https://openaccess-leidenuniv-nl.proxy.library.uu.nl/bitstream/handle/1887/52841/Masterscriptie J Bos - Locatie en bezit van hoofdelingen en eigenerfden in laatmiddeleeuws Friesland.pdf?sequence=1

Bucher, B., Falquet, G., Metral, C., & Lemmens, R. (2016). Enhancing the management of quality of VGI: contributions from context and task modelling. *European Handbook of Crowdsourced Geographic Information*, 131–142. https://doi.org/10.5334/bax.j

Casparie, W. A. (1993). The Bourtanger Moor: endurance and vulnerability of a raised bog system. In *Netherlands-Wetlands* (pp. 203–215). Dordrecht: Springer Netherlands.

https://doi.org/10.1007/978-94-011-2042-5_10

Casparie, W. A., & Molema, J. (2015). Het Middeleeuwse Veenontginnginslandschap bij Scheemda. *Palaeohistoria*, *32*, 271–289. Retrieved from https://ugp-rug-nl.proxy.library.uu.nl/Palaeohistoria/article/view/24905

Cooper, D., & Gregory, I. N. (2011). Mapping the English Lake District: A literary GIS. *Transactions of the Institute of British Geographers*, *36*(1), 89–108. https://doi.org/10.1111/j.1475-5661.2010.00405.x

Cowan, G. (1998). *Statistical Data Analysis*. *Oxford Science Publications* (Vol. 54). Clarendon Press. https://doi.org/10.1017/CBO9781107415324.004

Creative Research Systems. (2016). Sample Size Formulas. Retrieved December 4, 2018, from https://www.surveysystem.com/sample-size-formula.htm

De Vries, F., Hendriks, R., Kemmers, R., & Wolleswinkel, R. (2008). *Het veen verdwijnt uit Drenthe Het veen verdwijnt uit Drenthe Omvang, oorzaken en gevolgen*. Retrieved from www.boomblad.nl/rapportenservice

De Vries, G. (n.d.). Smildervenen. Retrieved October 22, 2018, from http://landschapsgeschiedenis.nl/deelgebieden/20-Smildervenen.html

Dempsey, C. (2011). Historical Geography and GIS. Retrieved October 17, 2018, from https://www.gislounge.com/historical-geography-and-gis/

European Commission. (2018). Natura2000. Retrieved October 25, 2018, from http://ec.europa.eu/environment/nature/natura2000/index_en.htm

Fryske Akademy. (n.d.). HISGIS. Retrieved October 24, 2018, from http://www.hisgis.nl/hisgis/over

Gabriela, D., Stefan, S., & Aurelian, B. (2010). Modern Techniques for Evaluation of Spatial Data Quality. *RevCAD – Journal of Geodesy and Cadastre*, (January 2009), 265–272. Retrieved from http://revcad.uab.ro/upload/18_162_Paper29_RevCAD09_2009.pdf

Geiger, C. P., & Von Lucke, J. (2012). Open government and (linked)(open)(government)(data). *JeDEM-EJournal of EDemocracy and Open Government*, *4*(2), 265–278. https://doi.org/10.29379/jedem.v4i2.143

Geonovum. (2016). *Metadata monitor NGR: Handleiding voor kwaliteitsverbetering metadata*. Retrieved from http://www.geonovum.nl/onderwerpen/metadata/documenten/kwaliteitsbeleid-en-norm-metadata-

Geonovum. (2017). Nederlands metadataprofiel op ISO 19115 geografie. Retrieved from https://www.geonovum.nl/geo-standaarden/metadata/nederlands-metadataprofiel-op-iso-19115-geografie

Gerding, M. A. W. (1995). *Vier Eeuwen Turfwinning: De verveningen in Groningen, Friesland, Drenthe en Overijssel tussen 1550 en 1950*. Landbouwuniversiteit Wageningen. Retrieved from https://library.wur.nl/WebQuery/wurpubs/fulltext/136833

Gerding, M. A. W. (2009). Het landschap van het “Wilde Wonen”. Hoogveenontginningen en heidevelden als woonoorden. In *Historia Agriculturae 33* (pp. 165–195).

Girres, J.-F., & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, *14*(4), 435–459. https://doi.org/10.1111/j.1467-9671.2010.01203.x

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial*

Statistics, *1*, 110–120. https://doi.org/10.1016/j.spasta.2012.03.002

Gregory, I. N. (2008). Different Places, Different Stories: Infant Mortality Decline in England and Wales, 1851–1911. *Annals of the Association of American Geographers*, *98*(4), 773–794. https://doi.org/10.1080/00045600802224406

Gregory, I. N., & Ell, P. S. (2006). Error-sensitive historical GIS: Identifying areal interpolation errors in time-series data. *International Journal of Geographical Information Science*, *20*(2), 135–152. https://doi.org/10.1080/13658810500399589

Gregory, I. N., & Healey, R. G. (2007). Historical GIS: structuring, mapping and analysing geographies of the past. *Progress in Human Geography*, *31*(5), 638–653. https://doi.org/10.1177/0309132507081495

Grift, J. (2015). *Veen in de Vallei*. Rijksuniversiteit Groningen. Retrieved from https://www.rug.nl/research/kenniscentrumlandschap/mscripties/masterscriptie_jeroen_grift_april_2015.pdf

Groenewoudt, B. J. (2009). An exhausted landscape. Medieval use of moors, mires and commons in the Eastern Netherlands (pp. 149–180). https://doi.org/10.1484/M.RURALIA-EB.3.1169

Hartog, H. M. (1884). Heide-ontginning. *De Economist*, *33*(2), 553–568.

Hennig, S., & Belgui, M. (2011). User-centric SDI: Addressing Users Requirements in Third- Generation SDI. The Example of Nature-SDIplus. *Geoforum Perspektiv*, *10*(20). https://doi.org/10.5278/OJS.PERSK..V10I20.448

Heywood, D. I., Cornelius, S., & Carver, S. (2011). *An introduction to geographical information systems*. Prentice Hall.

Homburg, C. J. (1991). Over de vorming van veen, het winnen van turf en de gevolgen voor ons land. *Grondboor & Hamer*, *45*(5/6), 154–161.

International Organization for Standardization. (2013). ISO 19157:2013 - Geographic information -- Data quality. Retrieved November 5, 2018, from https://www.iso.org/standard/32575.html

Jong, C. de. (2009). Het ontstaan van het Hooge Veene. *Grondboor & Hamer*, *63*(3/4), 75–79. Retrieved from http://natuurtijdschriften.nl/search?identifier=568313

Kekem, A. J. van, Hoogland, T., & Horst, J. B. F. van der. (2005). *Uitspoelingsgevoelige gronden op de kaart; werkwijze en resultaten*. Wageningen: Alterra. Retrieved from https://library.wur.nl/WebQuery/wurpubs/reports/338759

Knol, W. C., & Noordman, M. W. M. (2003). *De kadastrale kaart van 1832: digitale ontsluiting en landschapsecologische toepassingen*. Wageningen. Retrieved from https://library.wur.nl/WebQuery/wurpubs/fulltext/16439

Knowles, A. K. (2002). *Past time, past place : GIS for history*. ESRI Press.

Knowles, A. K. (2005). Emerging Trends in Historical GIS. *Historical Geography*, *33*, 7–13. Retrieved from https://scholar.googleusercontent.com/scholar?q=cache:duqblAILua0J:scholar.google.com/+historical+gis&hl=nl&as_sdt=0,5

Knowles, A. K. (2014, January 2). The contested nature of historical GIS. *International Journal of Geographical Information Science*. Taylor & Francis. https://doi.org/10.1080/13658816.2013.850696

Knowles, A. K., Roush, W., Abshere, C., Farrell, L., Feinberg, A., Humber, T., & Wirene, C. (2008). What could lee see at Gettysburg? *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship.*

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). Geographical Information Systems and Science. In *Geographical Information Systems and Science* (pp. 363–382). John Wiley & Sons. Retrieved from https://blackboard.utwente.nl/bbcswebdav/pid-1070868-dt-content-rid-2657960_2/courses/U17-GIMA-109/33_Geographical Information Systems and Science%281%29.pdf

Löwenberg, P. (2014). Neotropical region: a shapefile of Morrone's biogeographical regionalisation. *Zootaxa*, *3802*(2), 300. https://doi.org/10.11646/zootaxa.3802.2.12

Nijstad, F. (1982). Het kadaster van 1830. *Groniek*, *76*, 1–7.

Olah, B. (2009). Historical maps and their application in landscape ecological research. *Ekologia*, *28*(2), 143–151. https://doi.org/10.4149/ekol_2009_02_143

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211. https://doi.org/10.1145/505248.506010

Pleijter, M. (2004). *Veengronden en moerige gronden op de bodemkaart van Nederland anno 2003: onderzoek naar de afname van het areaal veengronden rondom schoonebeek*. Wageningen. Retrieved from https://library.wur.nl/WebQuery/wurpubs/427605

Reeskamp, A. S. (2013). *De Weerribben aan het einde van de vroege Middeleeuwen (800-1000 AD)*. Rijksuniversiteit Groningen. Retrieved from https://www.rug.nl/research/kenniscentrumlandschap/mscripties/ma_1464205_as_reeskamp.pdf

Rijksdienst voor Cultureel Erfgoed, & Ministerie van Onderwijs, C. en W. (n.d.). HISTLAND. Retrieved October 9, 2018, from https://data.overheid.nl/data/dataset/0da22798-aa17-4ef9-a35b-b54b3b9d3480

Rijksdienst voor het Cultureel Erfgoed. (n.d.). Beeldbank van de Rijksdienst voor het Cultureel Erfgoed. Retrieved December 4, 2018, from http://beeldbank.cultureelerfgoed.nl/alle-afbeeldingen/?q=&mode=gallery&view=horizontal&page=1&fq%5B%5D=search_s_head_collection:%22Kadastrale kaarten 1811-1832%22&reverse=0&sort=random%7B1543928534273%7D asc

Safe Software. (2019). FME: How FME works. Retrieved January 17, 2019, from https://www.safe.com/how-it-works/

Schouwenaars, J. M., Esselink, H., Lamers, L. P. ., & Molen, P. C. van der. (2002). *Ontwikkelingen en herstel van hoogveensystemen*. Wageningen. Retrieved from http://library.wur.nl/WebQuery/wurpubs/fulltext/85217

Sırrı Mara, S., Hadi Mara, H., Aktu, B., Emin Mara, E., & Yildiz, F. (2010). *Topological error correction of GIS vector data*. *International Journal of the Physical Sciences* (Vol. 5). Retrieved from http://www.academicjournals.org/IJPS

Spek, T. (2004). *Het Drentse esdorpenlandschap : een historisch-geografische studie*. Matrijs. Retrieved from https://library.wur.nl/WebQuery/wda/1733524

Stol, T. (1997). Het turfwinningslandschap. In *Het Nederlandse landschap. Een historisch-geografische benadering* (7th revised, pp. 47–56). Utrecht: Utrecht. Retrieved from http://dare.uva.nl/search?metis.record.id=134782

Storm, M. H., Knotters, M., Schuiling, C., & Clement, J. (2015). *Audit Basisregistratie Topografie 2014 :*

resultaten van de tweede wettelijk vereiste externe controle op de kwaliteit van de BRT. Wageningen: Alterra. Retrieved from https://library.wur.nl/WebQuery/wurpubs/reports/497852

Storm, M., Knotters, M., Schuiling, R., Smidt, R., van Dorland, G., & Wegman, R. (2018). *Wettelijke Audit Basisregistratie Topografie 2017*. https://doi.org/10.18174/450878

Thissen, P. H. M. (n.d.). De Nederlandse staat en heideontginning, 1795-1961. *Historisch Geografisch Tijdschrift*, 38–48.

Turner, R. (2002). Managing Quality. In *Gower Handbook of Project management*. Gower Publishing.

Unwin, D. J. (1996). GIS, spatial analysis and spatial statistics. *Progress in Human Geography*, *20*(4), 540–551. https://doi.org/10.1177/030913259602000408

Van Beek, R., Maas, G. J., & Van Den Berg, E. (2015). Home Turf: An interdisciplinary exploration of the long-term development, use and reclamation of raised bogs in the Netherlands. *Landscape History*. https://doi.org/10.1080/01433768.2015.1108024

Van Beek, R., Paulissen, M., & Quik, C. (n.d.). Home Turf Project. Retrieved October 21, 2018, from http://www.boglandscapes.eu/project.html

Van den Berg, S. (2018). *Quality assessment of professional and VGI geo-data in The Netherlands*. University Utrecht. Retrieved from https://dspace.library.uu.nl/handle/1874/366398

van der Burg, M. (2007). *Nederland onder Franse invloed. Cultuurtransfer en staatsvorming in de napoleontische tijd, 1799–1813*. University of Amsterdam.

Van Loenen, B. (2006). Developing geographic information infrastructure: The role of information policies. Delft University Press. Retrieved from https://repository.tudelft.nl/islandora/object/uuid:6781e9dd-1468-4b31-bf88-ee1fbfffe4a1?collection=research

Van Oort, P. (2006). *Spatial data quality: from description to application*. Wageningen University, Wageningen. https://doi.org/10.1038/nmat853

van Tussenbroek, G. (2018). Geografie van arm en rijk: Het kohier van de tiende penning van Amsterdam (1562) in GIS. *Tijdschrift Voor Historische Geografie*, *3*(4), 242–255.

van Zanden, J. L., & van Leeuwen, B. (2012). Persistent but not consistent: The growth of national income in Holland 1347–1807. *Explorations in Economic History*, *49*(2), 119–130. https://doi.org/10.1016/j.eeh.2011.11.002

Vandenbroucke, M., & Leloup, G. (2017). Het kadaster laat in zijn kaarten kijken. De vele onderzoeksmogelijkheden van het “primitief” kadaster. *Onderzoek*, 22–29.

Veldhorst, A. D. M. (1991). Het Nederlandse vroeg-19e-eeuwse kadaster als bron voor andersoortig onderzoek, een verkenning. *Historisch Geografisch Tijdschrift*, *9*(1), 8–27. Retrieved from https://thg.verloren.nl/_file/145021/Veldhorst_A.D.M._-_Het_Nederlandse_vroeg-19e-eeuwse_kadaster_als_bron_voor_andersoortig_onderzoek,_een_verkenning_-_jaargang_1991-1.pdf

Wang, X., & Yin, Z.-Y. (1997). Using GIS to assess the relationship between land use and water quality at a watershed level. *Environment International*, *23*(1), 103–114. https://doi.org/10.1016/S0160-4120(96)00081-5

Wolff, W. J. (1992). The End of a Tradition: 1000 Years of Embankment and Reclamation of Wetlands in the Netherlands. *Ambio*. SpringerRoyal Swedish Academy of Sciences.

https://doi.org/10.2307/4313944

Worst, D. (2012). *Agrarische veenontginningen in oostelijk Opsterland (900 - 1700 AD). Een interdisciplinair onderzoek naar de natuurlijke landschapsopbouw, de nederzettings- en ontginningsgeschiedenis en het agrarische landgebruik langs de boven- en middenloop van het Koningsdiep.* NetzoDruk. Retrieved from https://www.narcis.nl/publication/RecordID/oai:pure.knaw.nl:publications%2Fcd7001a9-df56-4203-8f6b-3f2d7302301d

Zagwijn, W. H. (1986). *Nederland in het holoceen*. Rijksgeologische Dienst. Retrieved from https://books.google.nl/books/about/Nederland_in_het_Holoceen.html?hl=nl&id=kI4oAQAAMAAJ&utm_source=gb-gplus-shareNederland

Appendix A: The sample used for the attribute accuracy assessment

OBJECTID of parcels in Drenthe			
233	37331	88418	138117
1003	39984	89439	141274
1212	40854	117214	142459
3867	43270	118127	142872
5845	43973	119056	144419
6645	46537	120264	146289
9514	46635	120732	149179
10157	46724	121839	154181
14132	48662	121887	155141
15119	48925	124097	157753
16456	52147	124589	159543
16503	52669	125065	163515
17001	78507	127116	164814
17318	78923	128643	166290
17409	79453	129136	166444
18527	79740	129580	166904
19104	79839	129878	167566
20664	80478	130752	168149
21215	80692	132653	168396
21875	81531	132762	170138
23984	81831	133138	176173
25902	84133	134027	177029
28447	85329	135456	177939
28753	85851	137303	182407
32147	86418	137523	182639

OBJECTID of parcels in Groningen			
961	56904	109999	166105
1021	57590	111463	168064
3566	58419	114748	170934
6273	59813	114999	172259
8356	62462	115270	174568
8669	62737	117125	177729
14685	66661	118106	186790
22612	69299	125163	187788
27856	70287	126091	189247
28679	74123	138784	190556
28686	78489	140403	192370
30711	80453	143693	193125
33229	84744	143712	195693
35299	84950	145951	198983
35621	86103	148660	199280
36727	90518	150171	199970
39022	92820	151043	200474
39301	96392	151723	202612
40233	98179	155125	203976
45618	98512	156001	207145
47116	98986	156700	207951
47700	101233	158962	212250
50168	104266	159153	213836
50352	105655	161149	214914
54963	107959	164837	215826

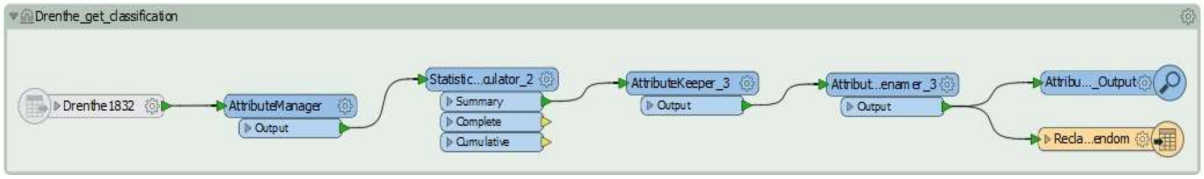
Appendix B: Format for attribute accuracy assessment

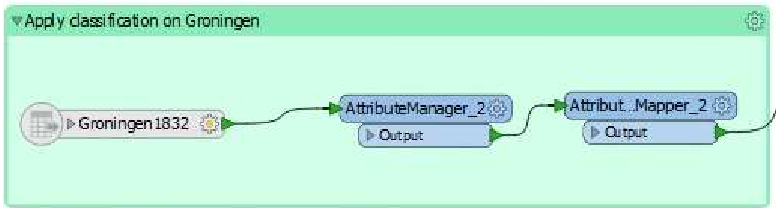
Attributes	Parcel A	Parcel B	Parcel C	Parcel XXX
object ID (ID)				
OAT (ID)				
Perceelnummer (ID)				
Naam				
Voornaam				
Beroep				
Woonplaats				
Soort_eigendom				
Inhoudsgrootte				
Klasse1				
Klasse2				
Klasse3				
Klasse4				
Klasse5				
Klasse_geb				
Tarief1				
Tarief2				
Tarief3				
Tarief4				
Tarief5				
Belastbaar inkomen				

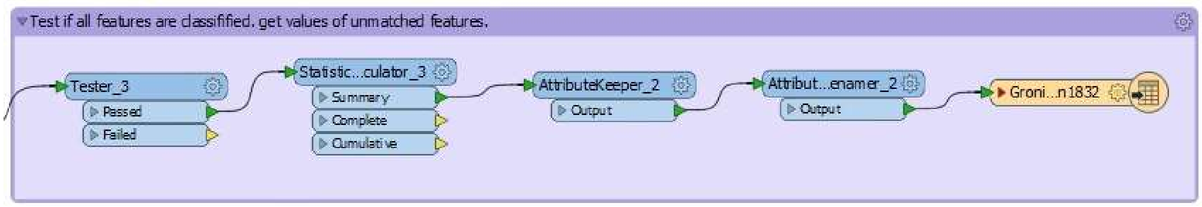
This format was used to assess the attribute accuracy. The red letters were the identifies and are not checked. The following fouls were noted:

1. The value 1 for wrongly digitised or typo errors.
2. The value 2 for missing values. Absent in the digital version, but present in original 1832 cadastre.
3. The value 3 for appearing values. Values in the digital version, but absent in the original 1832 cadastre.

Appendix C: FME scripts Reclassification





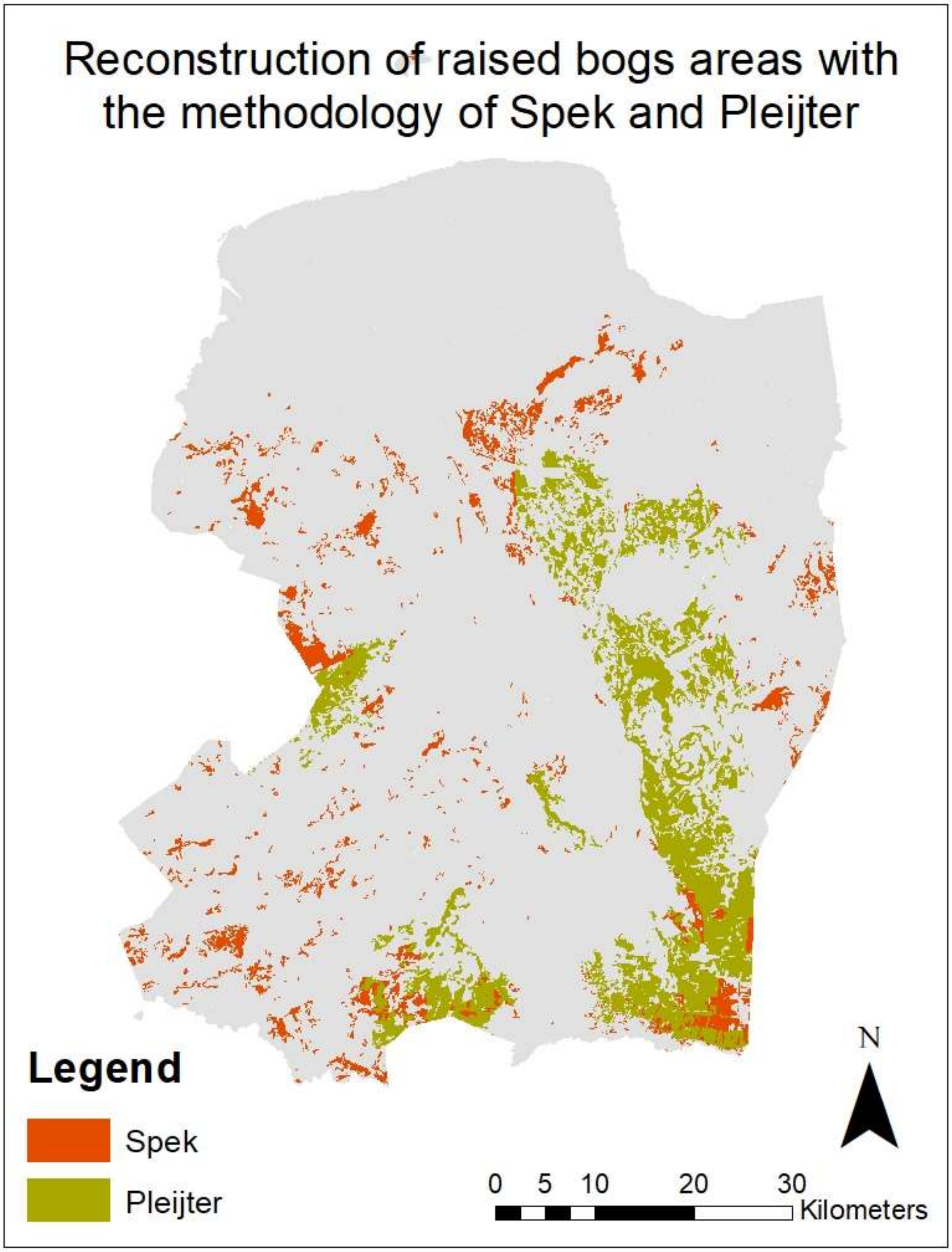


Appendix D: Classes of soil map used for reconstruction raised bogs

Legend	Description (in Dutch)
iWz	Moerige eerdgronden met een veenkoloniaal dek en een moerige tussenlaag op zand
iVz	Veengronden met een veenkoloniaal dek op zand zonder humuspodzol, beginnend ondieper dan 120 cm
iVp	Veengronden met een veenkoloniaal dek op zand met humuspodzol, beginnend ondieper dan 120 cm
iVc	Veengronden met een veenkoloniaal dek op zeggeveen, rietzeggeveen of moerasbosveen
iVs	Veengronden met een veenkoloniaal dek op veenmosveen
iWp	Moerige podzolgronden met een veenkoloniaal dek en een moerige tussenlaag
AP	Petgaten
Avo	Veen in ontginning
aVp	Madeveengronden op zand met humuspodzol, beginnend ondieper dan 120 cm
aVs	Madeveengronden op veenmosveen
vWp	Moerige podzolgronden met een moerige bovengrond
zVp	Meerveengronden op zand met humuspodzol, beginnend ondieper dan 120 cm
pVs	Weideveengronden op veenmosveen
Vp	Vlierveengronden op zand met humuspodzol, beginnend ondieper dan 120 cm
Vs	Vlierveengronden op veenmosveen
zVs	Meerveengronden op veenmosveen
zWp	Moerige podzolgronden met een humushoudend zanddek en een moerige tussenlaag

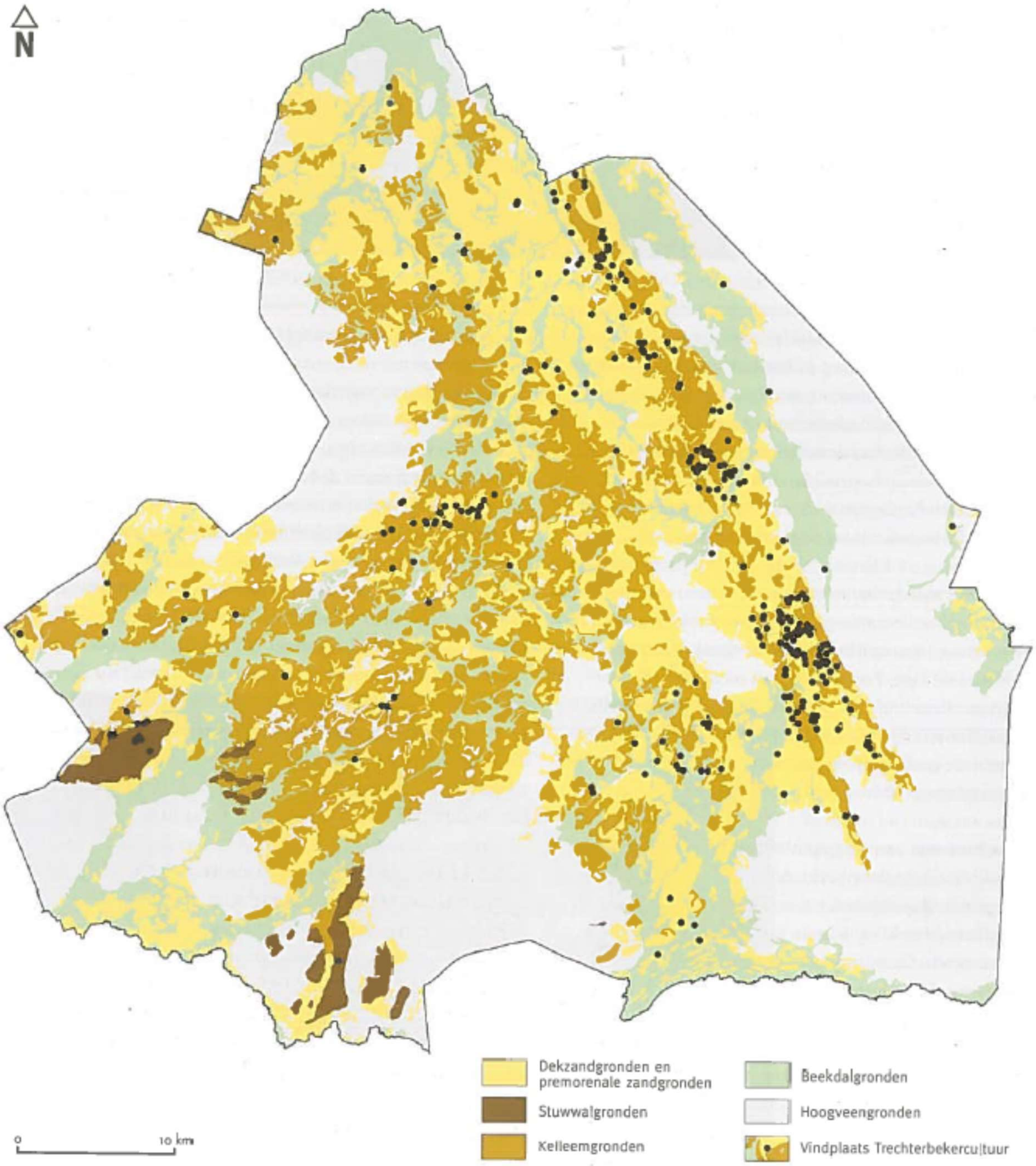
Legend and Description (in Dutch) of the soil classes in the Dutch soil map (Alterra, 2006)*. based on Pleijter* (2004) *and Spek* (2004)*.*

Appendix E: Reconstruction of raised bogs in the study area with methodology of Spek and Pleijter



Reconstruction of the raised bogs with the soil map and the methodology of Pleijter (2004) and Spek (2004)

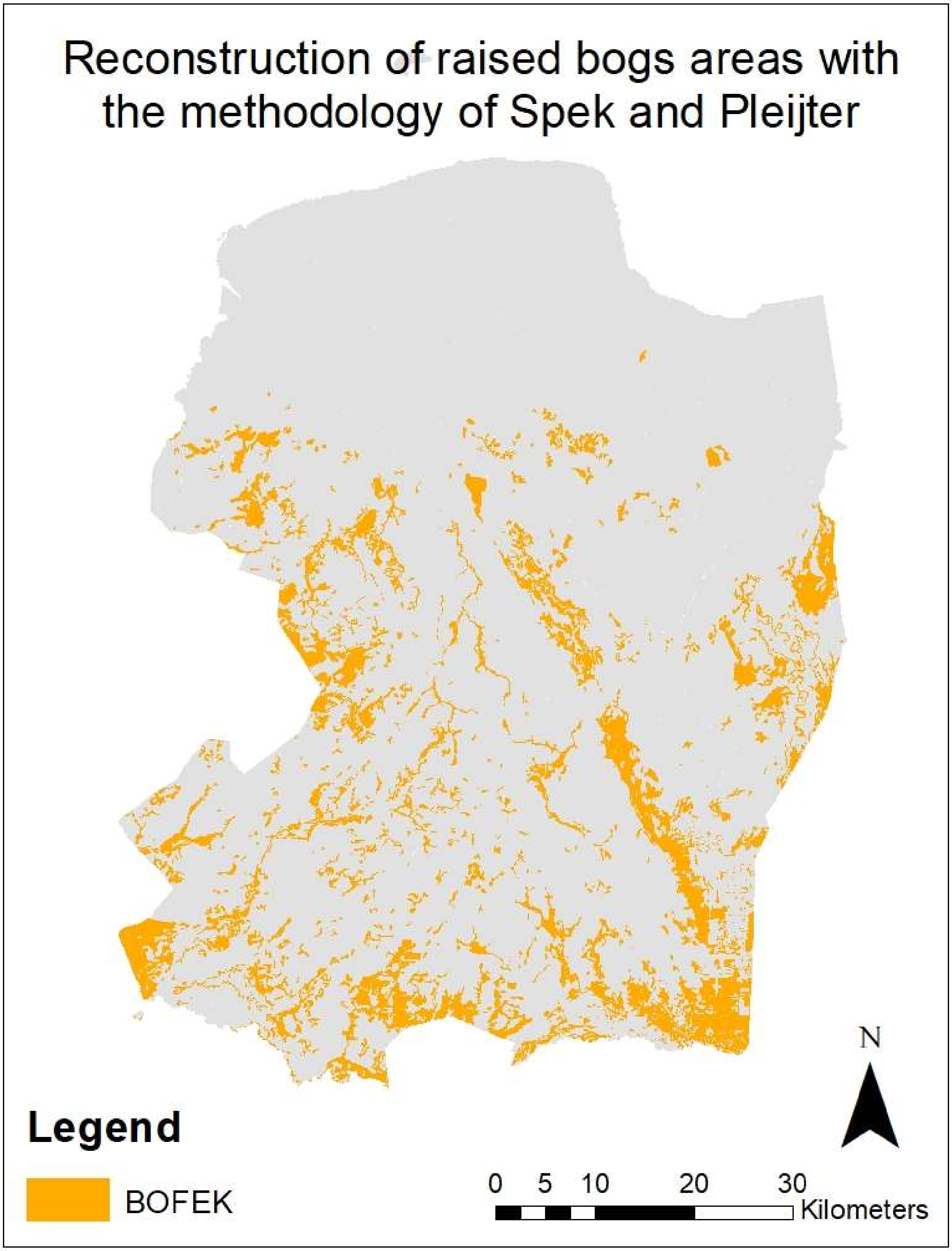
Appendix F: Reconstruction of Drenthe by Spek (2004)



Appendix G: The classes included in the BOFEK methodology

Number of the class	Description (in Dutch)
103	Kleiarme moerige bovengrond op veen met binnen 120 cm-mv. vaak een zandondergrond
104	Kleiarme moerige bovengrond op veen met binnen 120 cm-mv. een zandondergrond met leem
107	Oligotroof veen tot dieper dan 120 cm-mv.
108	Veen op zandondergrond binnen 120 cm-mv.
109	Zanddek of veenkoloniaal dek op mesotroof veen en een zandondergrond binnen 120 cm-mv.
110	Zanddek of veenkoloniaal dek op oligotroof veen met een zandondergrond binnen 120 cm-mv.
205	Zanddek op moerige tussenlaag op zandondergrond
206	Zanddek op moerige tussenlaag op zandondergrond met keileem

Appendix H: Reconstruction of raised bogs using the BOFEK methodology



Appendix I: Peat colonies in HISTLAND