

iSEGMound – a Reproducible Workflow for Mound Detection in LiDAR-derived DTMs

Agnes Schneider

Philipps-Universität Marburg

16.09.2021

Abstract

TEST TEST TEST TEST

Contents

1	Introduction	2
2	Exploitation of Automated Analysis Methods to detect burial mounds and structurally similar archaeological objects: a survey	6
3	The Necessity of a Paradigm Shift in Archaeological Remote Sensing practice by means of Tool-driven Automated Analysis	20
4	Applied Mound Detection	23
5	Results of the iSEGMound Workflow	24
6	Discussion	29
7	Conclusion	30

1 Introduction

The use of quantitative methods in archaeology reach back to the end of the 19th century, but it was only after the middle of the 20th century that they became computer-based. The first applications of statistical-mathematical methods using computers revolutionized the handling of spatial and quantitative archaeological data (e.g. Goldmann 1979) and also the view on archaeology itself (e.g. Clark 1968 and almost a decade later Hodder and Orton 1976, Ihm and Zimmermann 1978 to begin with). This quantitative analytical view preceded and prepared archaeologists for the advent of the public availability of soon high resolution digital datasets which revolutionized and broadened Aerial Archaeology, drawing upon the expertise of Satellite Remote Sensing and branching into a new discipline: Archaeological Remote Sensing. These new data sources, including geophysical platforms, led to the need to handle - in archaeological terms - big spatial data, thus to the specialized use of GIS platforms and the development and application of new methods like predictive modelling (van Leusen – Kamermans 2005; Kamermans et al. 2009). The sophistication of airborne platforms, sensors and imaging technologies, like LiDAR, hyper-spectral imagery and drone derived multispectral and hyper-spectral imagery (Agapiou – Lysandrou 2015; Luo et al. 2019) facilitated the diversification of the toolset of Archaeological Remote Sensing. New technical developments continuously push specialists to look for, borrow and adapt new methods to analyse in archaeological sense big data, which first lead to automating tasks such as georeferencing or applying the same function on multiple rasters (Raun and Patterson 2018, 70). Soon this lead to more complex automation – that is automating whole workflows and also analysis. Thus in contrast to other disciplines, Archaeological Remote Sensing and Archaeological Science itself was quite late in adopting and applying automated methods and thus, Automated Archaeological Remote Sensing is still in its infancy (Opitz and Herrmann 2018, 30). Although accordingly, the expression ‘automated analysis’ is still controversial, it is nonetheless used throughout this Master’s thesis. It has to be emphasized that ‘automation’ does not stand for completely automated workflows. It was chosen as a short phrase to address workflows with at least partly – or mostly – automated elements, because no one advocates ‘automatic archaeology’ (Cowley 2012, 6). Automated analysis, often also called semi-automated analysis, ‘simply’ means that a part, often a major part of a specific workflow, is automated, meaning computer aided. Critical voices about predictive modelling in GIS which can be seen as the precursor and actual starting point of automated analysis were very fittingly summarized by Wheatley 2004: predictive modelling is dehumanising, anti-historical and substitutes human actions with mathematical equations (Wheatley 2004). This distrust is reflected in the criticism towards automated analysis methods in the Archaeological and Archaeological Remote Sensing community, claiming, that archaeological projects are always locally specialized (Parcak 2009, 120) and that there is no generalized automation method that can be used location independently, which can locate atypical archaeological objects without producing a quite large miss-rate (Casana 2014, Casana 2020, S93). Taking these points into account: why only use one type of automated analysis method (Davis 2019, 5)? Large-scale landscapes feature lots of different landscape forms and archaeological features which need to be addressed in different ways. It is evident that it is not possible to detect everything with one analysis method or algorithm, due to the variability of the archaeological record. Other considerations include reflections on Automated Archaeological Remote Sensing detecting round and square shapes over and over and when it will move to something more useful that actually resembles archaeological objects (Rog Palmer in personal discussion and several AARGnews Editorials and contributions, e.g. AARGnews 62, 61)? It is also somehow self-evident, that although pre-trained and known object classes are going to be detected, new and unexpected and also not detected objects always have to be expected. Thus expecting to detect unexpected or unknown object classes with supervised learning/automation is misleading and completely unrealistic (see a similar discussion about the use of predictive modelling in Wheatley 2004, 3.2.3). It is a fatal move to expect automation to be the magic trick, when it is not: it should be more seen as an extension of the archaeologist’s toolbox. Because automated analysis is not to replace manual data evaluation which, based on the personal experience of the operator and them being human, can be prone to missing knowledge or error, or even archaeologists completely, but to be complementary. Automated analysis should be followed by field observation and identification for assessment, which on the other hand also can be biased for the same reason as manual data evaluation (Cowley 2012, Bennett et al 2014, 899). All in all automated analysis is to ease and facilitate the amount of work archaeologists are facing with the increasing amount of data and different data types to analyse and evaluate. As Davis (2020a, 3) emphasizes: automated analysis is precise and manual analysis is

accurate. Thus both methods combined can help archaeologists to penetrate a hidden level of any dataset and landscape, which is invisible or hardly visible even for the trained eye. Taking it further, reproducible analysis (Marwick 2017, Rokem et al. 2017, Marwick et al. 2018, Calero Valdez 2020) can help to facilitate automated analyses and can serve as control for human individuality: because archaeologists do see different things in the same data set and unconsciously see what they are familiar with and what they know (Cowley 2012, 7), very similarly to an algorithm trained to find a certain class of object. Human detection cannot be reproduced but automated analysis can: via documented, reproducible workflows. In order to achieve this goal open-source scripting languages like R, Python or open-source platforms like Google Engine should be used, to be able to repeat, reproduce and even replicate workflows to facilitate their use by other scholars. When writing code, the semantic syntax has to be clear and consistent. This also makes sure that the ontology and the semantic syntax will be the same when the code is used by a different operator on a different dataset – in contrary to manual analysis, where independent, subjective manual operators would define features or objects differently – thus arriving to the same or at least similar bias and errors, of course depending on the dataset. This argument is investigated more thoroughly in chapter III and the expressions reproducible and replicable are discussed in more detail.

Being entirely different disciplines, there is a semantic gap between the ontologies of Computer Vision, Image Analysis and Remote Sensing on one hand and Archaeology on the other hand – especially because the archaeological record itself is fragmented, multifaceted and poses ontological problems in itself for the interoperability of projects or databases, conducted by different operators with possibly different metalanguage. Although the aim of this thesis is not the creation of an ontological and semantic framework and/or a metalanguage for Automated Archaeological Remote Sensing, these points have to be discussed because they affect the way how automation can be harmonized with Archaeological Remote Sensing.

This fundamental difference is reflected in the fact, that a distinction has to be made between ‘object’ and ‘feature’ in *sensu stricto* automation, based on termini technici from Computer Vision, on which ground ‘object’ is referred to real-world entities in remote sensing images and ‘feature’ to elements of an image and of an object (Traviglia et al. 2016, 12; Lambers et al 2019, 2), in contrary to the every-day use of the expression ‘archaeological feature’ at an excavation site or in reports. Reproducible code and workflows enable to conduct the same ‘procedure’ in a controlled environment and thus possible semantic problems which can result from how different operators see archaeological objects and features can be detected, distinguished and solved. Thus the accurate expert knowledge of manual analysis can be integrated in the precision and consistency of computational semantics (Davis 2020a, Fig 1.).

On the other hand, manual analysis can learn a lot from the systematic precision of reproducible code: defining variables, consistent nomenclature of objects/ forms/ features; creating functions: setting the relation of objects/ forms/ features. This needs the consistency of formalized, straightforward and clear definitions (also Magnini and Bettineschi 2019, 13, Davis 2020a). Thus ontology, semantics and metadata are very important tools which enable us to connect and share code, databases and research. By shared ontologies (knowledge representation), codified metalanguage protocols and rule-sets, not only would the transferability of expert knowledge be substantially eased between Archaeological disciplines and Automated Archaeological Remote Sensing, but also between its different subdisciplines, like Template Matching-, Geometric knowledge-, GeOBIA-, Machine Learning- based analyses (this distinction is explained in chapter 2).

So then, after all, how to semantically address the varying, unpredictable, fragmented, multifaceted and diachronic archaeological record (monuments and artefacts) which can be detected in remote sensing data with methods of Computer Vision and Image Analysis? Remote sensing data offers much more: they are diachronic time capsules, records of palimpsest landscapes, archaeological, micro- and macro-topographical, geomorphological and also recent, anthropomorphic traces (Magnini and Bettineschi 2019, 12) in a multi-dimensional space. Thus a common ground has to be created (a shared diachronic formalized ontology) to fill the semantic gap between the metalanguage of human perception, the archaeological record (which is manifold, imbalanced and changing in time (conceptually and physically) and Computer Vision/Image Analysis. And a way has to be found to apply this new diachronic formalized ontology onto real world scenarios (using Automated Archaeological Remote Sensing). Real world (archaeological) scenarios are often extremely complex and thus robust simplification and formalization has to be executed to break down complexity and bring the domains in discussion together. Sevara and Pregesbauer (2014) already pioneered a conceptual framework in 2014, which found resonance and the formulation of a Diachronic Semantic Model

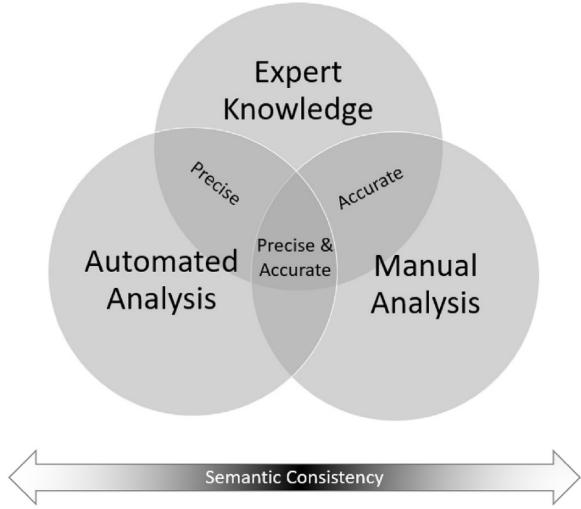


Figure 1: Chart showing the advantages (approaching Semantic Consistency) when Expert Knowledge, Manual Analysis and Automated analysis are combined. Source: Davis 2020a, Fig 1. (#fig:Figure 1)

(DSM) in Magnini and Bettineschi 2019, which was successfully applied by the latter on a case study in 2021 (Magnini and Bettineschi 2021). These fundamental studies are the first steps towards integrating Automated Archaeological Remote Sensing in the nervous system of Archaeological Science.

As concluding thoughts let a quote from Quintus et al. (2017, 1) state: “*many archaeological professionals who might have an interest in lidar-derived products do not have the technical experience to modify or create AFE (automated feature extraction) techniques for particular regions or environments.*” This should not impact the scientific methods one chooses. With open-source, reproducible (regularly updated and version controlled) and code, with clear replicable workflows and published data and study (at least the manuscript in a preprint repository) (Figure 3), every remote sensing archaeological professional should be able to tap into the possibilities of Automated Analysis. Thus open access and open source software should be used, training data sets and workflows released to be able to learn from each other and to progress, building on the experience of each other and not to have to start from scratch with every new project conducted.

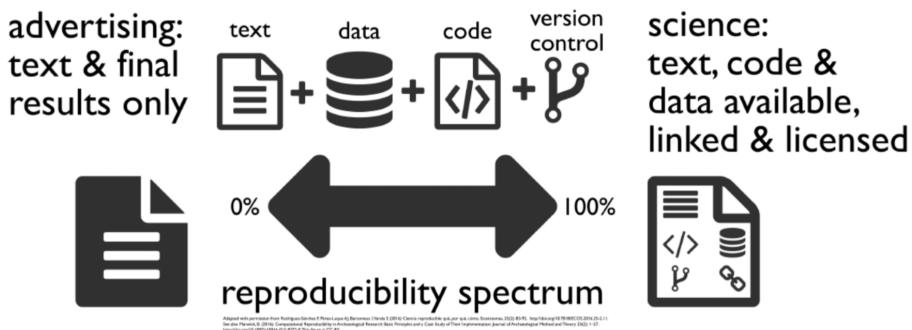


Figure 2: The difference between articles behind paywall and reproducible (and replicable) studies. Source: Marwick et al. 2017, Figure 1. (#fig:Figure 2)

Thus we not only need to compare the accuracy and the true and false positive rate of different automated

methods (Trier et al. 2018, Table 1) and the methods themselves (Davis et al. 2019), but we need to make a compendium of these methods with the key elements mentioned above. Combined with this, a baseline or best practice for the different methods should be defined that can be utilized by beginners. Thus open-access code, workflow and protocol, published (at least training) data sets and best practices are needed as reference and base line for other studies, stored in an open-source database or platform which can support open-source computer languages.

Different observations made above and questions based on these have led to the objective of this thesis. Firstly: based on scientific literature, on what grounds lies the dispute of automation in Archaeological Remote Sensing and how can it be relieved? Is the experience and knowledge gained transferable to other studies? How many scientific studies on burial mound detection worked transparently and retraceable (reproducible if not replicable) in any step of their workflow? Can experiences and knowledge from previous studies be used to create a reproducible and replicable workflow for burial mound detection?

Thus the aim of this Master's thesis is to emphasize the applicability and usefulness of automation in dealing with big datasets in Archaeological Remote Sensing and the need for reproducible and replicable workflows and studies. The case study for this thesis is the detection of burial mounds in LiDAR datasets in R, to substantiate this point of view.

To approach these questions accordingly, after the discussion of Automated Archaeological Remote Sensing and its reception in Archaeological Science was discussed in Chapter I, the Introduction. In the next chapter (II) all published studies whose goal is to detect burial mounds or mound-like structures are discussed in terms of used methods, software, accessibility of code and data. It is also elaborated why burial mounds are chosen as Objects of Interest and which methods will be applied and why. Chapter III provides the toolbox for the thesis to implement the desiderata and requirements discussed in Chapter I and II. Chapter IV presents a reproducible workflow for the detection of burial mounds in LiDAR data. Chapter V (Discussion) and Chapter VI (Outlook) completes the endeavour of this Master's thesis.

2 Exploitation of Automated Analysis Methods to detect burial mounds and structurally similar archaeological objects: a survey

The analysis carried out and communicated in this chapter reviews all published studies which are aimed at the detection of burial mound and mound-like structures in order to define a baseline to place present research. Numerous reviews of methods and applications used in Archaeological Remote Sensing have been carried out (e.g. Leisz 2013, Agapiou and Lysandrou 2015; Davis 2018, Luo et al. 2019, Fiorucci et al. 2020): some broad and general, some more specific. The word survey was chosen specifically, because it was not a systematic search of databases as various ‘classical’ reviews did (such as Agapiou and Lysandrou 2015, Raun and Patterson 2018, Magnini and Bettineschi 2019 and Davis 2020a & b). This survey is building on the experience of aforementioned systematic reviews (which show very diversified nomenclature and heterogeneous terms) and the scientific literature was collected based on these reviews, Researchgate, Academia and mainly considering open-access publications (if possible). As a starting point all published studies (at least available to the author via University access; number of studies in August 2021: 96) have been collected which deal with the automation of Archaeological Remote Sensing data set (practical studies which use any automation methods and analysis) but only those are discussed, which are concerned with the detection of burial mounds and structurally similar archaeological objects: altogether 31 publications. Burial mounds have been chosen as Objects of Interest for the review because they are relatively simple structures and a very common funerary custom throughout human history. Similarly structured archaeological objects like tell mounds or other monumental earthworks of circular form can be found in very different areas and time periods and are widely investigated and thus deliver good examples of automated analysis methods of these archaeological objects. The first archaeological objects to automatically detect were tell mounds and only then and with the dawn of the use of LiDAR data came burial mounds into the focus of research. While the early research carried on tell settlements can be connected to Ur & Menze, the monumental earthworks and mound shell rings can also be allocated to two research groups: Freeland et al. and Davis, Lipo & Sanger. The detection of burial mounds themselves shows more variability in the investigators (Figure 3). A complete list of references (which is by far not conclusive and is to be supplemented) is included as a supplement at the end of this thesis.

```
##  
## Attache Paket: 'dplyr'  
  
## Die folgenden Objekte sind maskiert von 'package:stats':  
##  
##     filter, lag  
  
## Die folgenden Objekte sind maskiert von 'package:base':  
##  
##     intersect, setdiff, setequal, union
```

The main points considered in this survey were chosen based on the aforementioned reviews, which were conducted looking at citation indexes of different Remote Sensing methods in scientific periodicals (Agapiou and Lysandrou 2015, Raun and Patterson 2018, Luo et al. 2019, Davis 2020b), Institutional Affiliations of researchers (Agapiou and Lysansrou 2015, Raun and Patterson 2018), the citation network (Raun and Patterson 2018), OBIA by geographic region (Davis 2018), the developments and limitations of OBIA methods (Davis 2018), research goal using OBIA (Davis 2018), the development of different passive and active Remote Sensing methods (Luo et al. 2019), the scale of ArchaeOBIA applications (Magnini and Bettineschi 2019), datasets and methods applied in Archaeological Remote Sensing (Fiorucci et al. 2020), different parameters, thresholds, methods and accuracy used in the automated detection of mound features (Davis 2020a) and geometric disparity of Remote Sensing methods (Davis 2020b). These systematic reviews stimulated questions such as: How can we learn from previous studies? How transparent is the workflow and if described, which variables and parameters (rule sets) were used? Which software was used? How reproducible or even replicable is the workflow in the studies?

Objects of Interest in Automated Archaeological Remote Sensing between 2006 and 2021

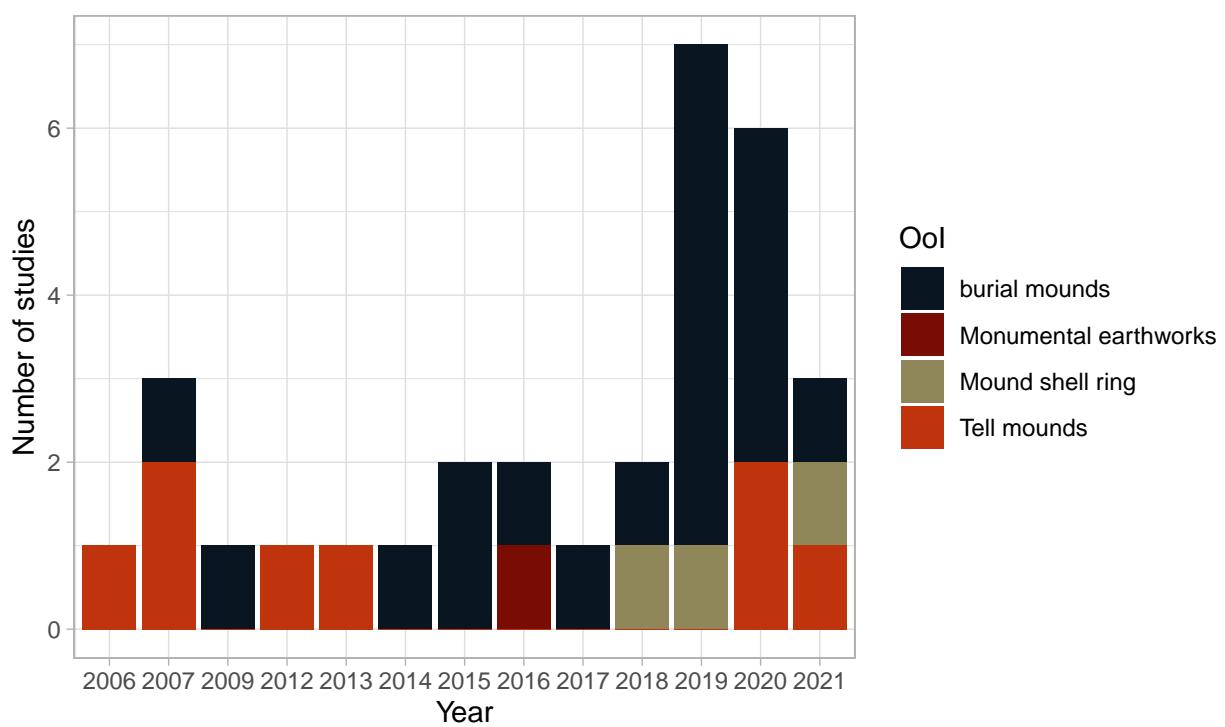


Figure 3: The Objects of Interests investigated by year between 2006 and 2021.
(#fig:Figure 3)

Based on these questions the points investigated in this Master's thesis are:

- (i) Location (only in Supplement 1)
- (ii) RS (Remote Sensing) data
- (iii) Methods
- (iv) Detail of methods (only in Supplement 1)
- (v) Variables/morphometric parameters
- (vi) OoI (Objects of Interest)
- (vii) Scale
- (viii) Software
- (ix) Access

The point “Location” (i) was documented, because different geographical and landscape conditions ask for different methods and objects can appear geographically and culturally diverse, which can explain the chosen method. The point “RS data” (ii) documents which type of remote sensing data was used and also its resolution if known. The point “Methods” (iii) organizes the studies in five main categories (Template Matching, Geometric Knowledge-based, GeOBIA-based and Machine Learning-based, Deep Learning; see Figure 4), based on Cheng and Han 2016 (also see Lambers et al. 2019 and Roffler 2020), instead of juxtaposing pixel-based and object-based methods. This differentiation points out the different object detection methods based on the approach to the dataset. Still, for the differentiation of Machine Learning-based methods, the expression pixel-based is going to be used for the classifier training-based methods, also because several studies use it (Sevara et al. 2016, Niculită 2020b).

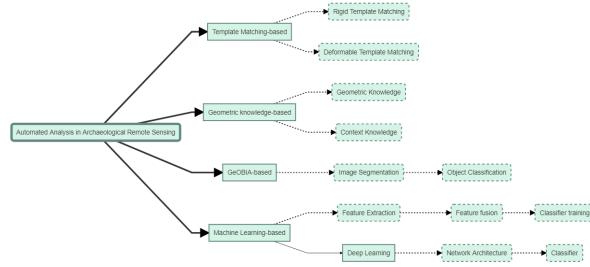


Figure 4: Taxonomy of Automated Analysis in Archaeological Remote Sensing, based on Cheng and Han 2016.

(#fig:Figure 4)

The point “Details of method” (iv) explains the workflow of the study in a nutshell, if comprehensible. Point v, “Variables/Morphometric parameters” highlights the variables and or morphometric parameters used in the study. “OOI” (vi) are the objects investigated. The point “Scale” (vii) implies the geographical range of the study, if it was applied to a broader region than the area on which the method was developed. The last two points (“Software” - viii, and “Access” - ix) investigated give us hints about the accessibility of the dataset and the code. Especially in earlier studies, information about the software used is not mentioned and sometimes can only be guessed. In the case of “Access”, apart from the information about accessible data and code it was also documented if there is a workflow (a comprehensible sequence of the steps of the workflow) or a flowchart (a very generalized workflow without clear steps) or if any of the equations used was published and or explained. It was also documented if any supplementary information or document was attached to the study which can guide to a better understanding for a possible reproduction of the study. Although all these points have been investigated, for practical reasons mainly the points (ii) RS (Remote Sensing) data, (iii) Method/s, (iv) Detail of method/s, (v) Variables/morphometric parameters, (vi) OoI/Object(s) of Interest, (viii) Software and (ix) Access is going to be elaborated in depth in this and the next chapter.

It is essential to discuss the basic traits of the used methods (iii) in short, concentrating on their application in Archaeological Remote Sensing. It has to be stressed, that still no semantic lingua franca (between Archaeological Remote Sensing and Computer Vision and Remote Sensing Methods) exists when talking about automated methods in Archaeological Remote Sensing, thus methods are addressed in different ways (or most commonly in a very simplified way) and it is not always possible to determine which exact method has been utilised (recently it is improving, but it depends where the authors put the focus of their work). First, when shortly discussing the basic traits of the methods and their use in Archaeological Remote Sensing, it is only focussed on the use of the different methods (what is used, where and when) since 2006, when tell mounds were first investigated using automated analysis methods. In the second part of this chapter the way of their use is investigated and analysed how the methods are used.

Template Matching-based methods utilize a template of the Object of Interest to be detected, which is then statistically matched to the input image. Two general directions can be distinguished: Rigid Template Matching, which requires a precise template which gets problematic when it comes to intra-class shape and size variations. Deformable Template Matching on the other hand can handle either free-form deformable templates or parametric deformable templates. (Cheng and Han 2016, 13-14). In the literature investigated Template Matching was used seven times for the detection of burial mounds, mound shell rings and tell mounds (De Boer 2007, Kramer 2015, Trier et al. 2015, Davis et al. 2018, Raum 2019, Davis et al. 2019, GholamReza and Malian 2021). Apart from Kramer 2015, Rigid Template Matching was preferred (Table 1). Template Matching itself is a method which is being revisited from time to time (Figure 5).

```
## [1] "Year"    "Method"   "Freq"
```

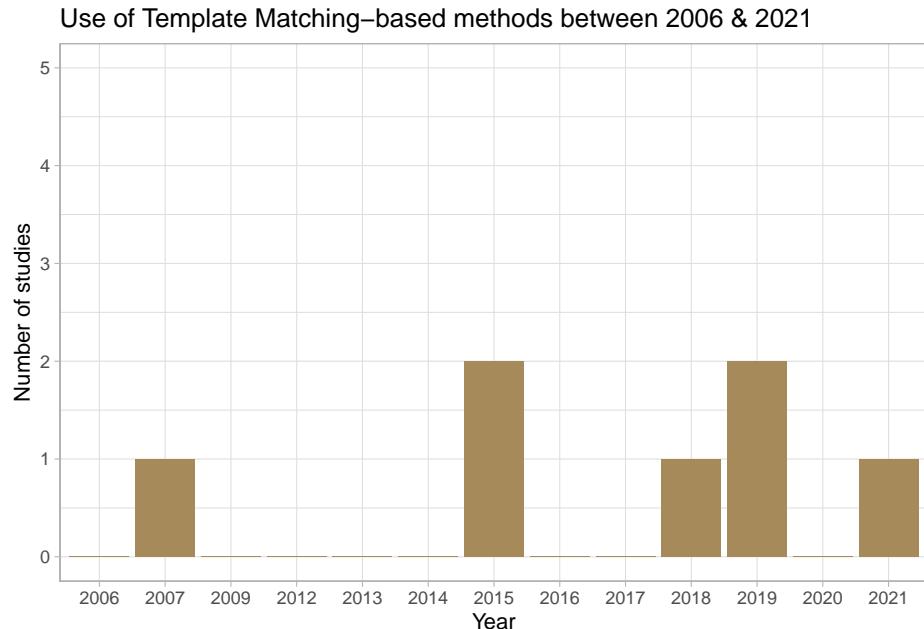


Figure 5: The Objects of Interests investigated by year between 2006 and 2021.
(#fig:Figure 5)

Geometric knowledge-based analysis works with specialized solutions for specific problems, applying rules based and established on knowledge of the regions of interest and its context. It uses either encoded prior geometric knowledge of the generic specificity of the Object of Interest and then e.g. applies hand-crafted mathematical morphometry rules for object extraction (morphometric derivatives such as Slope, Aspect, Curvature, Local Relief Model, vegetation indices etc.). Context knowledge based analysis uses knowledge about the relationship of the Object of Interest and the area it should be separated from (e.g. filters,

textural analysis) (Cheng and Han 2016, 15). In the case of Automated Archeological Remote Sensing the two Geometric knowledge-based analysis approaches are often used in combination and form the data preparation step (Figure 6). Geometric knowledge based object analysis is only occasionally used per se as an automated analysis method for the detection of burial mounds, monumental earthworks and tell mounds: Riley 2009, Freeland et al. 2016, Rom et al. 2020. (Lately) it is only occasional, that Geometric knowledge-based object analysis is not used as data preparation method (De Boer 2007, Caspari et al. 2014, Trier et al. 2015, Raun 2019, Caspari and Crespo 2019, Kazimi et al. 2019a, Kazimi et al. 2019b). In some cases, Geometric knowledge based analysis is included in the future work (compare works by Kazimi et. al 2019 vs. Kazimi et al. 2020)(Table 1.).

```
## [1] "Year"    "Method"   "Freq"
```

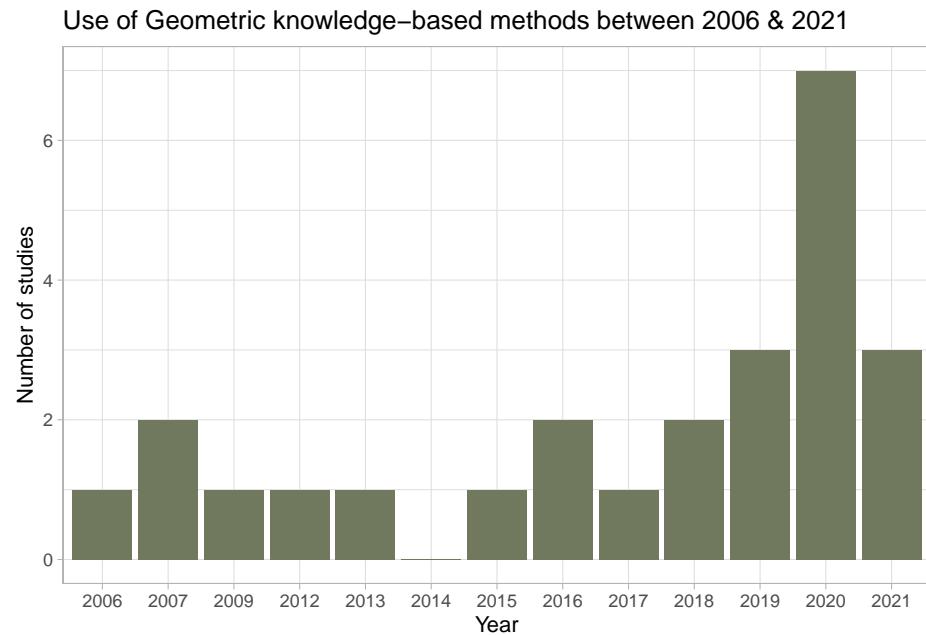


Figure 6: The use of Geometric knowledge-based methods distributed by year between 2006 and 2021. (#fig:Figure 6)

Geographical Object-based Image Analysis (GeOBIA) dates back to the late 1970's (Blaschke et al. 2014), but it was only around 2000 that Object-based Images Analysis became widely used in Remote Sensing studies due to the availability of high resolution Satellite data. This induced a paradigm shift not only in Remote Sensing but generally in GI Science, hence its new name: GeOBIA (Hay and Castilla 2008). GeOBIA operates in two steps: images are divided into small segments, which are defined by the homogeneity of specific morphometric (shape, size, orientation), spectral, textural, context and neighborhood parameters (Hay – Castilla 2008) and are then grouped together to meaningful homogeneous object candidates (the segmentation step), which are then classified by specific extracted object criteria in question (the feature extraction and classification step, Blaschke et al. 2014, 186, Cheng and Han 2016, 15, Hossain and Chen 2019, 115) or filtered by a rule-set. Various types of segmentation methods exist (as also their categorization: Blaschke 2010, Blaschke et al. 2014, Kumar et al. 2014). Lately Hossain and Chen 2019 investigated the different segmentation methods from a Remote Sensing point of view, but here only the two main Segmentation methods used in Automated Archaeological Remote Sensing studies are described: Watershed Segmentation (or Transformation) (Nicuță 2020a,b) and Multiresolution Segmentation (Kramer 2015, Sevara et al. 2016, Freeland et al. 2016, Davis et al. 2018, Davis et al. 2019, Meyer-Heß et al. 2019, Sărăsan et al. 2020), the first being an Edge-Based Segmentation method, the latter a Region-Based Segmentation method. For an in-depth analysis see Hossain and Chen 2019.

Edge-Based Segmentation techniques are ‘top-down’ methods: first they locate edges in the image and then use contouring algorithms to close them. Edges are regarded as boundaries between objects where pixel properties are abruptly changing (Hossain and Chen 2019, 117). ‘Watershed Segmentation’ (implemented in open source software e.g. in SAGA or in the ForestTools package in R) simulates real-life flooding. It first transforms the image into a gradient (grey-scale), then identifies objects with clear segment boundaries, only to then create (fill) objects, for which it is also called a Region-Growing algorithm (Hossain and Chen 2019, 117, Table 1; Figure 7).

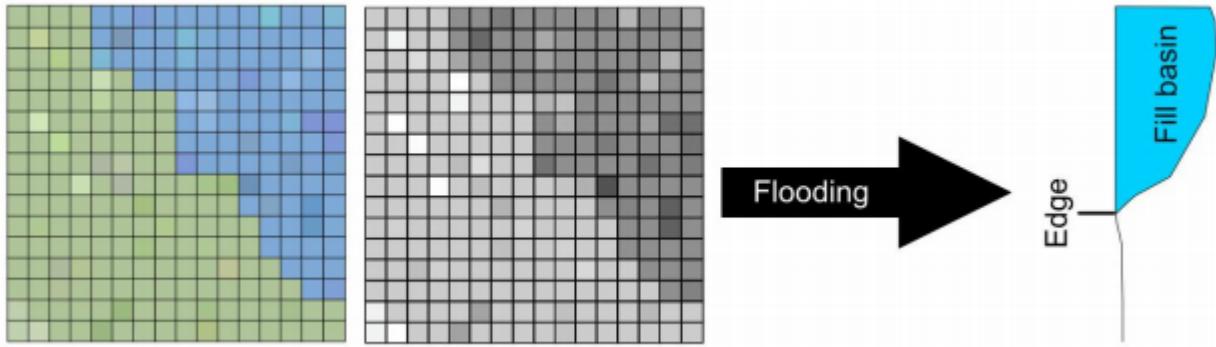


Figure 7: Operating principle of ‘Watershed Segmentation’. Roffler 2020, 33.
(#fig:Figure 7)

Region-Based Segmentation techniques start from the complete opposite: they begin with an initial segmentation of the whole image (thus also called ‘bottom-up methods’) based on a specific rule-set. Regions are generated in two completely different ways: either by splitting the image into homogeneous regions based on inhomogeneity (region-splitting and then merging) or by region-growing from a so-called seed-region based on homogeneity (region-growing and then merging, Hossain and Chen 2019, 119). ‘Multiresolution Segmentation’ (Baatz and Schäpe 2000, for which mainly eCognition is used in Automated Archaeological Remote Sensing) is a region-merging hierarchical segmentation method which starts with one pixel (seed) and applies pairwise merging of segments to build up hierarchical levels. The merging – or clustering (using local rule sets) is repeated (on multiple levels), until an object is recognized (Hossain and Chen 2019, 119, Roffler 2020, 33-34; Figure 8).

With regard to (Automated) Archaeological Remote Sensing, GeOBIA is sometimes also called archaeOBIA (Lamotte and Masson 2016), which points to the fact, that (Automated) Archaeological Remote Sensing is in need of a completely different semantic ontology and rule sets than which is needed for customary GeOBIA methods used in Remote Sensing. Still recently the challenge in Remote Sensing has been to define segmentation parameters/rule sets which can be transferred to other images (Cheng and Han 2015, 16) – something archeOBIA is also struggling with (see Table 1; note the different variables used in the different studies). Burial mounds, monumental earthworks and mound shell rings have been investigated using GeOBIA methods nine times since 2015 (Figure 9). Apart from one case (Nicuță 2020: Watershed Segmentation, carried out with SAGA in R), Multiresolution Segmentation was applied (Table 1), almost exclusively using eCognition, (former Definiens), a software developed for and with the evolution of GeOBIA and thus is generally considered “the software” for GeOBIA (Blaschke 2010, Fig 4, Hossain and Chen 2019, Fig 1.), which is clearly reflected in this survey.

```
## [1] "Year"    "Method"   "Freq"
```

Machine Learning-based methods are considered a subfield of Artificial Intelligence. Machine learning automates statistical methods to learn from input data either supervised, unsupervised or semi-supervised. Automated Archaeological Remote Sensing mainly utilizes supervised methods. *Pixel-based Image Analysis*, an image classification method, has been developed in the early 1970s for the digital analysis of Landsat

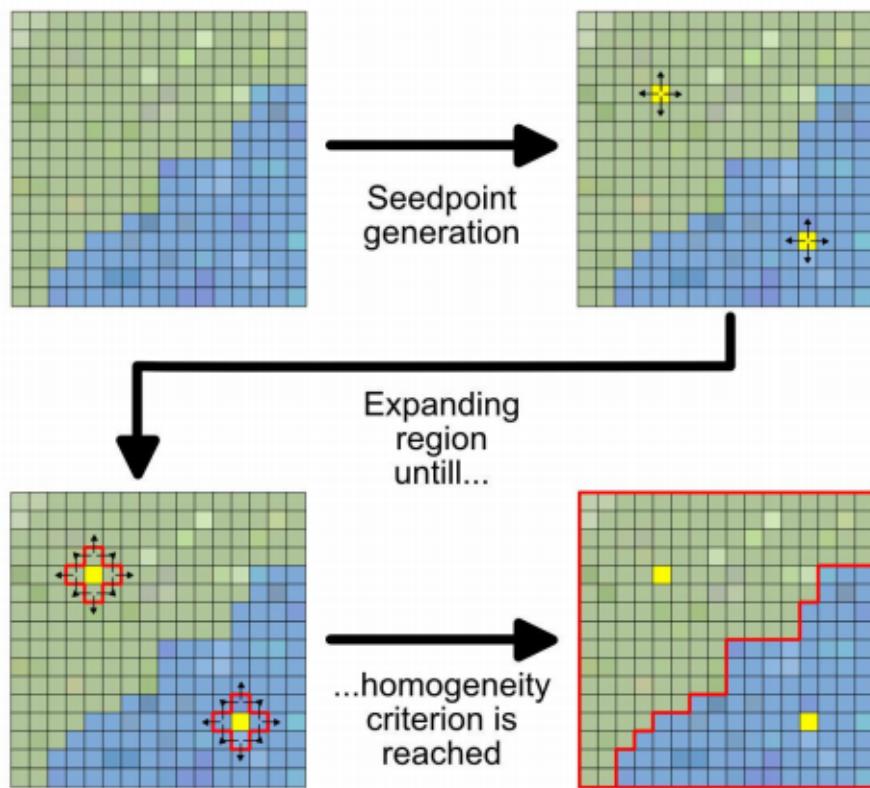


Figure 8: Operating principle of 'Region Growing Segmentation'. Roffler 2020, 34.
(#fig:Figure 8)

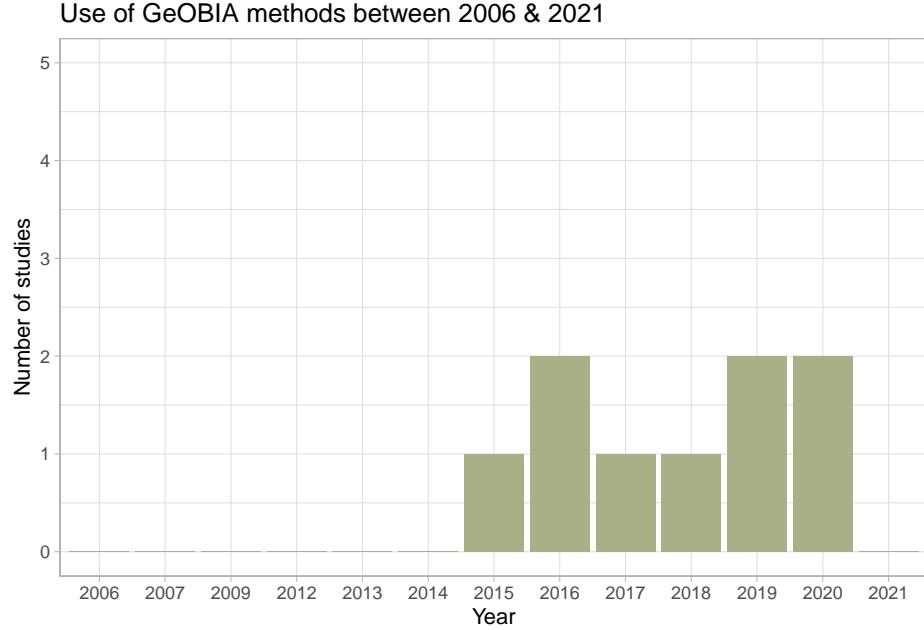


Figure 9: The use of GeOBIA methods distributed by year between 2006 and 2021.
 (#fig:Figure 9)

Multispectral Scanner Systems (Phiri and Morgenroth 2017, 9) and is (still) widely distributed in Remote Sensing research, especially in land-use and land-cover classification. In contrast to GeOBIA, Pixel-based Image Analysis approaches an image on pixel basis, which are classified into different categories based on the information they carry (usually one variable). Given that the second step of GeOBIA can be a classification of the segmented image using variables of the image-objects best describing the Objects of Interest, GeOBIA is sometimes also seen as a form of Machine Learning (Davis 2018, 1). Random forest classifiers are supervised learning algorithms which consist of an ensemble of decision trees. Each (unrelated) decision tree is trained using a random subset of the training data. Each of these trees will give a prediction for a datapoint. Then, the prediction of all decision trees is averaged by a majority vote to a final prediction (Figure 9). The independence of the different decision trees increases the accuracy of the prediction and also eliminates problems that can be caused by outliers in the dataset and works also well with small datasets, because of the facts just explained. These effects can be enhanced by resampling techniques and parameter tuning (Kuhn and Johnson 2016).

Regarding the automated detection of burial mounds and similar structures, Pixel-based classification was the first method used in 2006 and has been more or less revisited since (Figure 11). When looking closely at the specific algorithms used, it is not surprising why the Random Forest algorithm was exploited in most Pixel-based Image Analysis studies (Menze et al. 2006, 2007, Menze and Ur 2007, 2012, 2013, Kramer 2015, Guyot et al. 2018, Orengo et al. 2020, Nicușă 2020, Davis et al. 2021) and Mahalanobis Distance (Trier et al. 2015, Severa et al. 2016) and Support Vector Machine Classifiers (Caspari and Crespo 2019) in less.

```
## [1] "Year"    "Method"   "Freq"
```

It was already suggested that with the development of remote sensing sensors and resulting new, higher resolution datasets the need for new analysis methods is constantly stimulated. Starting with pixel-wise analysis and followed by the object-level addressing of remote sensing imagery, lately an even bigger semantic step was taken: the analysis on the scene-level, which can be seen as the next paradigm shift (Cheng and Han 2016, Cheng et al. 2020, 2, Fig. 2). The complex semantic structures of very high resolution images are addressed by *Deep Learning*, which is also a sub-field of Machine Learning. In contrast to Pixel- and Object-based Image Analysis, Deep Learning Algorithms consist of multiple stacked hierarchical layers (network

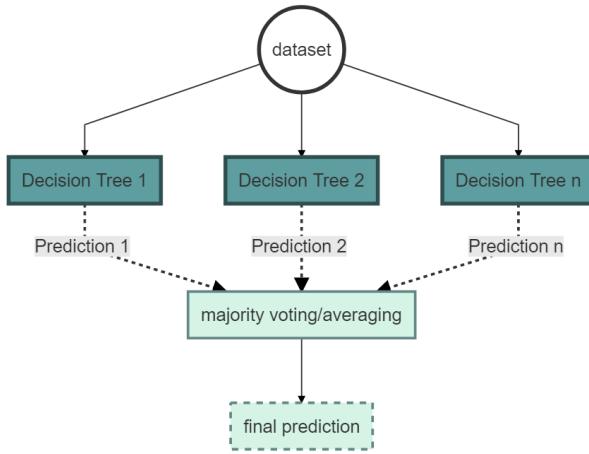


Figure 10: Schematized modus operandi of the Random Forest algorithm.
 (#fig:Figure 10)

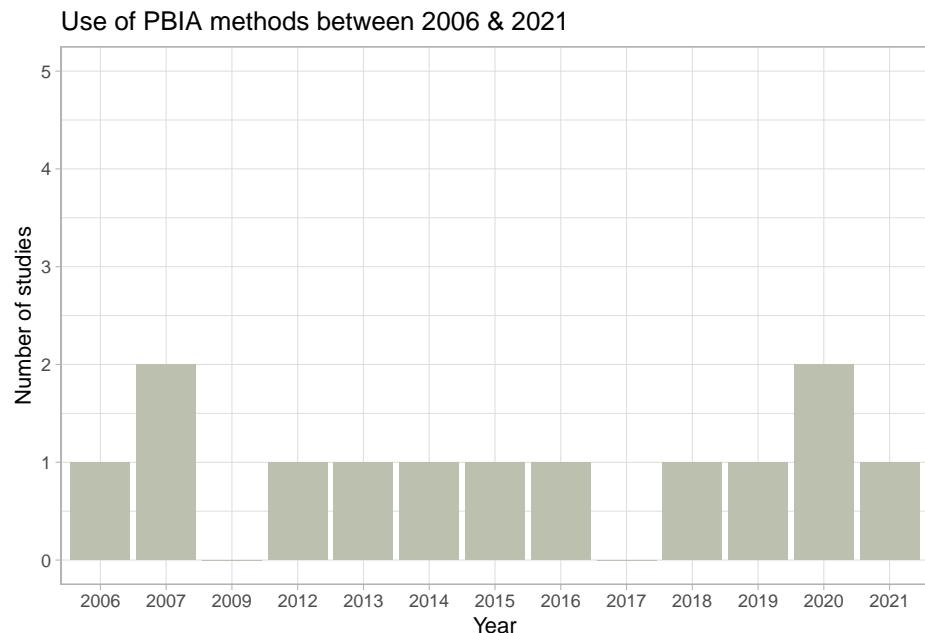


Figure 11: The use of Pixel-based Image Analysis methods distributed by year between 2006 and 2021.
 (#fig:Figure 11)

architectures) which can handle complexity and abstraction. In the case of the identification of burial mounds and mound like structures Deep Learning is only in use since 2019 (Figure 12).

```
## [1] "Year"   "Method" "Freq"
```

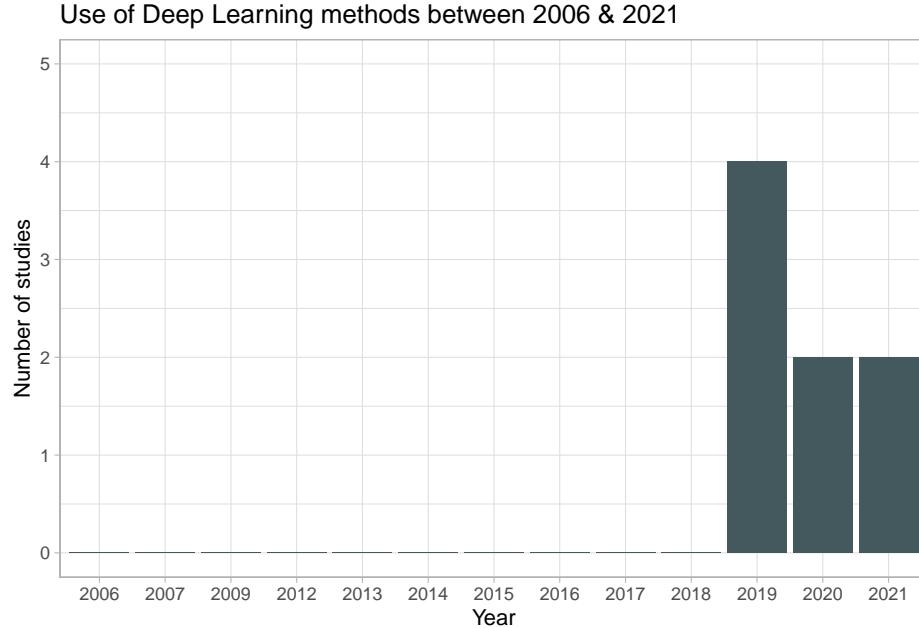


Figure 12: The use of Deep Learning methods distributed by year between 2006 and 2021.
(#fig:Figure 12)

To summarize the use of automated analysis methods to detect burial mounds and mound-like structures (Figure 13), it can be established that the first method used was Pixel-based Image analysis (Menze et al. 2006), followed by Template Matching (de Boer 2007). Geometric knowledge-based analysis was used as an autonomous method only by Riley 2009, Freeland et al. 2016, Rom et al. 2020, but as already concluded it is more often than not incorporated in workflows as data preparation method. GeOBIA was first used in 2015 and remained a very effective analysis method, until recently when Deep Learning became widespread (including Semantic and Instance Segmentation).

Several studies compare different analysis methods or combine them. Freeland et al. 2016 and Davis et al. 2019 compare Geometric-knowledge-based analysis (iMound/Inverse Stochastic Depression Analysis) to GeOBIA, with success (this is also reflected in the fact that the original algorithm iMound by Freeland et al. 2016 was reused by Davis et al. 2019 and Rom et al. 2020). Template Matching has been compared to GeOBIA (Kramer 2015, Davis et al. 2019: including Geometric knowledge-based method) and Pixel-based Image Analysis to GeOBIA (Sevara et al. 2016) and to Deep Learning (Caspari and Crespo 2019).

Although LiDAR data has been available for some time (see De Boer 2007, Riley 2009), it was only after 2010 that it made its way into everyday archaeological research, including various visualization methods (as Geometric knowledge-based analysis and data preparation method), making LiDAR visualisations a self-evident step for any archaeological project using ALS data (see also Kokalj and Hesse 2017). It is mainly from 2015, when ALS data started to dominate and revolutionize Automated Analysis methods in Archaeological Remote Sensing and provoking new approaches, such as GeOBIA (Figure 14). Since the diffusion of LiDAR data, Satellite Imagery is mainly used in large-scale studies (Caspari and Crespo 2019, Orengo et al. 2020). Studies which require high resolution data and utilize LiDAR quickly shifted to Deep Learning solutions. Simultaneously also UAV data finds its way into the general data repertoire of Automated Archaeological Analysis (Sărăşan et al 2020).

Methods used in Automated Archaeological Remote Sensing
between 2006 and 2021

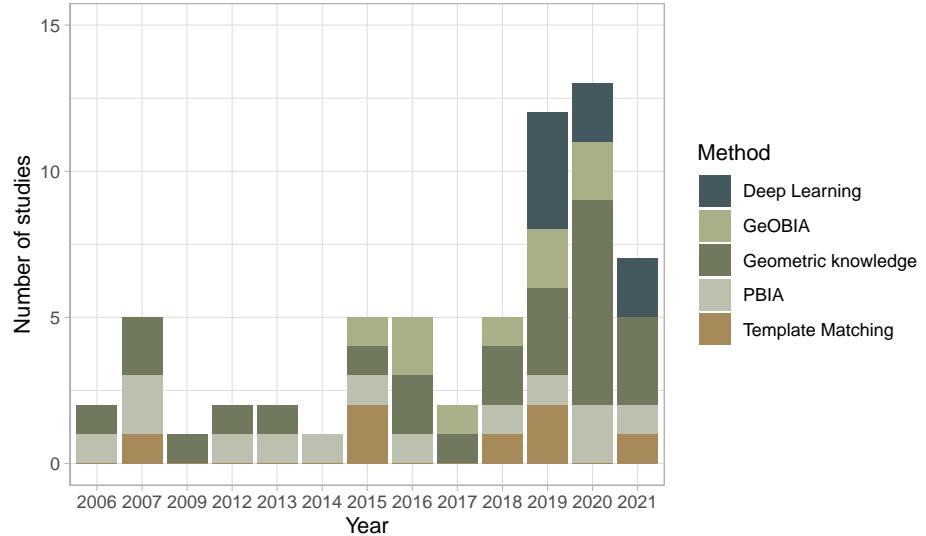


Figure 13: The use of different Image Analysis methods to detect mounds and mound-like objects distributed by year between 2006 and 2021. Note: this represents the number of methods used, not the number of studies. (#fig:Figure 13)

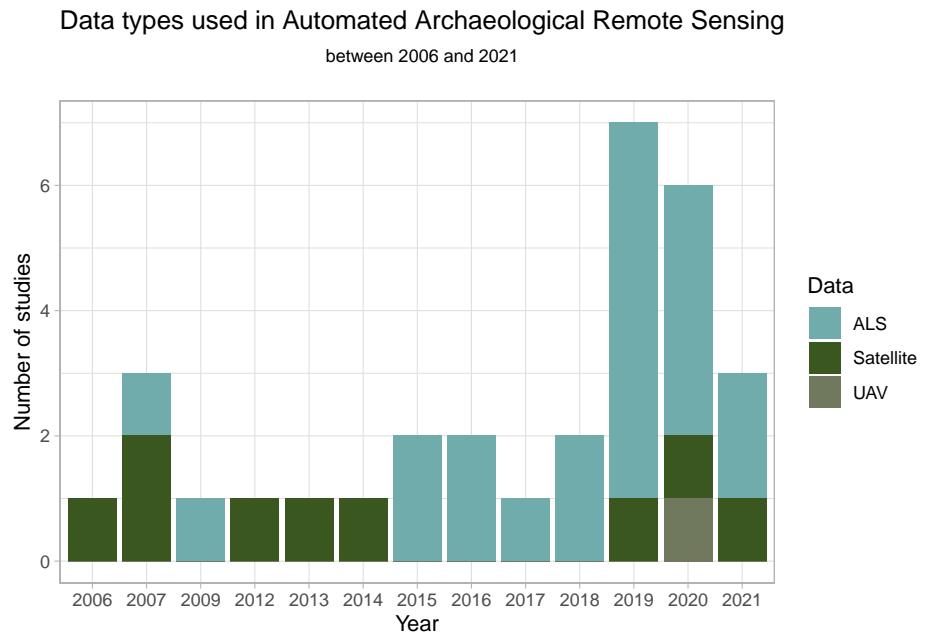


Figure 14: The use of different data types used to detect mounds and mound-like objects distributed by year between 2006 and 2021. (#fig:Figure 14)

When taking the points ‘Software’ and ‘Access’ into account, it has to be expressed that it is only a recent phenomenon that software and computation details have to be disclosed when publishing a study. Even though lately many journals require data and code availability statements, only a few studies provide openly accessible code and data: Orengo et al. 2020, and Niculiă 2020. In the first case the data and the code are available using Google Earth Engine. In the latter case, the regulations of the local Cultural Heritage Management authorities require a signed form through the author of the study to be able to use the DEM on which the study was based on - nonetheless it can be accessed. Although with restrictions. Inspecting the information about the methodology details of the studies, nine cases have been identified (Figure 15): from not available (n/a) to workflow & code & data, many constellations can be observed. In many cases the workflow was published, in some cases only a flowchart. In this thesis as a workflow the clear explanation of the methodology in a chart form is defined (with a big probability of reproducibility if the observer knew the exact tools and software used or if those were made clear). As a flowchart on the other hand, a part of a workflow or a very schematized workflow was identified where only the main steps were arranged in chart form, thus making it impossible to trace back the specific steps and to reproduce the workflow or any part of it. Thus the only reproducible study, which published a really detailed workflow, the code and made the dataset - although restricted by certain rules - available is Niculiă 2020.

Access to any parts of studies investigated

between 2006 and 2021

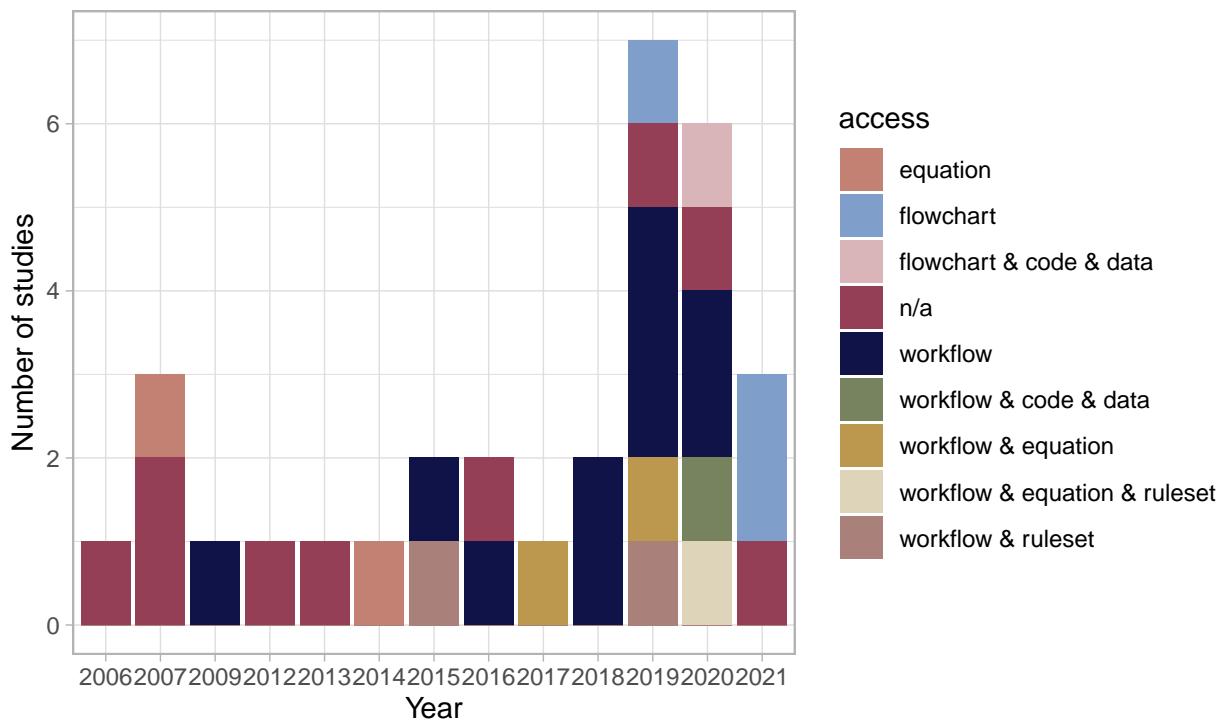


Figure 15: Access to any parts of studies used to detect mounds and mound-like objects distributed by year between 2006 and 2021.

(#fig:Figure 15)

As already noted, eCognition is considered “the” software for GeOBIA (Hossain - Chen 2019, 122) thus it is evident that many will choose this simpler solution. At the same time, it must be emphasised that the first analyses were carried out in R (Menze et al. 2006, Menze et al. 2007, Menze and Ur 2007), i.e. with open source software.. Using proprietary software not only marginalises researchers and institutions who/which can’t afford rather expensive software, but it is also hard to comprehend the exact algorithm behind the

software, and to understand to be able to apply it in another domain. It is also important to understand why certain tools or algorithms worked or did not work? This can only be done by creating reproducible workflows with open source software. In 2019 and 2020 more than 60% of the studies were conducted with FOSS (Free and Open Source Software) software. This number is only increasing (Figure 16).

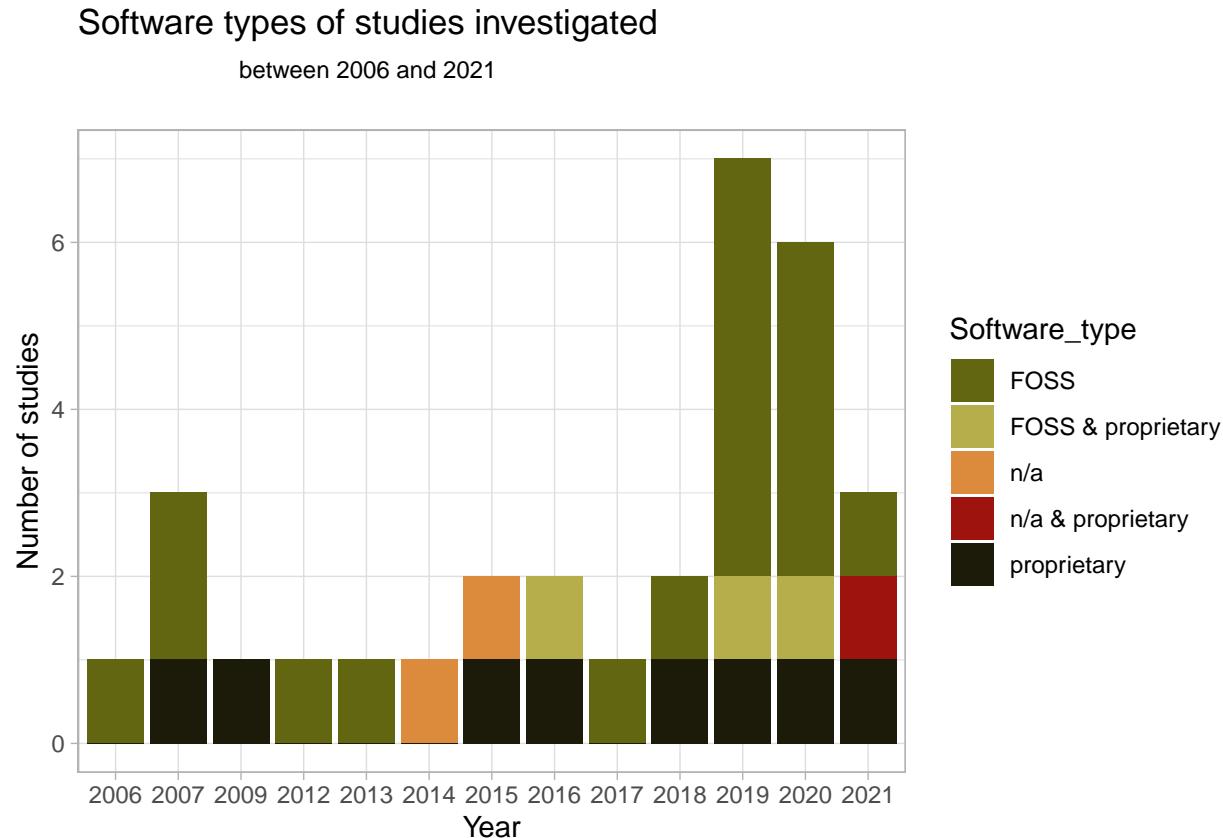


Figure 16: Software types of studies used to detect mounds and mound-like objects distributed by year between 2006 and 2021.
(#fig:Figure 16)

This survey served the purpose to elaborate the development of Automated Analysis in Archaeological Remote Sensing, mainly with regard to the applied methods, the used software and access to information about the workflow. The aim was to show how little reproducible research has been published with regard to automated analyses and how much there is to be done in the future. Also, automated analysis in Archaeological Remote Sensing is being carried out on different scales with different algorithms, specialized to different research questions in mainly research contexts. This is also a reason why many critical voices see automated analysis methods as pastimes and not really something operational (see many AARGnews Editorials). To give way for the next chapter we can conclude that the results of automated analyses are compelling, but many studies start basically from the beginning, because when research is published without accompanying software, workflow, data and documented steps, it is a challenge and time consuming to understand, verify, expand and eventually surpass that research (Marwick 2017, 425). This is one reason why many research projects start from the beginning, with a new idea instead of building on the previous knowledge. Another point is that automated analysis methods in Archaeological Remote Sensing are still in their infancy (Opitz and Herrmann 2018, 30) and best practices have not yet been established. Nonetheless it has to be stated that automated analysis itself is not a goal, but a method, a means to an end to further research and thus reproducible or even replicable best practices would help shift the focus on further development of methods than on continuous recommencements.

Chapter II discussed the use of automated analysis methods in Archaeological Remote Sensing for the detection of burial mounds and structurally similar archaeological objects. It was demonstrated that only 2 out of 31 studies have disclosed workflow, code and data and are thus reproducible (see chapter III for explanation). As a reproducible study in the R environment Niculiță 2020 can be taken as the best example. It was also noted that even without the availability of code, there are workflows which are illustrated and also explained clearly enough that they can be reproduced, which on the other hand can take some time but is still possible to do when having enough experience with spatial tools. On this basis it was decided that the Geometric knowledge-based workflow ‘iMound’ (established by Freeland et al. 2016 and reused by Davis et al. 2019 and Rom et al. 2020) is going to be utilized in R to detect burial mounds in LiDAR data in this Master’s thesis to create a reproducible workflow. During the implementation of the workflow it became clear that on the basis of the geomorphometry of the Train Area it was necessary to include another method to be able to effectively detect burial mounds (this is discussed in depth in Chapter IV). Thus, in addition to ‘iMound’ also the reproducible workflow of Niculiță 2020, thus a GeOBIA method was used in this Master’s thesis.

3 The Necessity of a Paradigm Shift in Archaeological Remote Sensing practice by means of Tool-driven Automated Analysis

In Chapter II the arguments of reproducibility and replicability have been addressed regarding the accessibility of data, code and workflow of the investigated studies. It was demonstrated that only 2 out of 31 studies have disclosed all three elements and are reproducible. It is important to discuss the exact definition of both expressions. In this Master's thesis mainly the definitions of Marwick 2017 are followed. Reproducibility means that code and workflow of an analysis reproduces all results and figures in the respective publication from its original raw data. Replicability stands for the replication of the workflow (from data collection to data analysis) onto a new dataset. It was mentioned in the introduction that optimal reproducible open science would incorporate version controlled open code, clear workflows and open datasets or at least training datasets.

So then, how is the best way to implement and guarantee reproducibility? Marwick 2017 outlined four general principles of reproducible research: i) Data and code provenance, sharing and archiving (analogue to artifact provenience) ii) Scripted analyses (for reproducibility and ideally replicability) iii) Version control (transparency of decision points of any research and possibility of collaboration) iv) Self-contained computational environments (sharing the computational environment of the analysis for replicability).

Pushing forward along this road, Marwick et al. 2018 propose R-packages as research compendia to organize transparent and reproducible research to approach the general principles laid out in 2017. Research compendia should organize files according to conventions: separate data, method and output (the latter being disposable because they can be reproduced) and define which exact computational environment was used for the analysis. Building an R package around a research project automatically forces one to keep the guidelines and conventions of package building which also provides quality control mechanisms. Research compendia come in many forms and complexities, as Marwick et al. 2018 discusses. The author of this Master's thesis experimented with a basic form of reproducible research compendium when analysing the finds of an Iron Age cemetery in Hungary using multivariate statistical analysis (Schneider 2020). The work was done basically in an R project with a fixed project structure which was uploaded in a github repository (https://github.com/keltoskytoi/Multivariate_Statistics_Szentloerinc). The repository contains an .Rproj file with the settings for the project, a folder called DATA for the raw data to be analyzed and not to be changed during the study, two .R files with the code used for the analysis in the study, a README.md file with basic information of the paper, thus the title and where to get it. An earlier study and research compendium from 2015, but already in a very mature form by Ben Marwick already shows the many of the elements of an optimal research compendium (listed in Figure 12), organized as an R package with the published compendium on figshare for the study Clarkson et al. 2015 which is also completely replicable: <https://github.com/benmarwick/1989-excavation-report-Madjedbebe>; https://figshare.com/articles/software/1989_excavation_report_Madjebebe/1297059 A recent use of a research compendium is Schmid 2019 with the compendium published at OSF. This study is also completely replicable: (<https://github.com/nevrome/cultrans.bronzeageburials.article2019>, <https://osf.io/b6np2/>) Based on the reproducibility of these studies and on the personal experience and motivation mentioned above, it was decided to create a research compendium for the Master's thesis.

The optimal outline for a complex research compendium in an R environment is reproduced from Marwick et al. 2018, Figure 4:

This optimal research compendium outline is implemented in the R package rrtools ('reproducible research tools', <https://github.com/benmarwick/rrtools>) by Ben Marwick, which helps to write reproducible research papers, setting up the structure propagated in Marwick et al 2018. Although this Master's thesis cannot give place to all of these desired steps and tools, it shall provide a reproducible workflow for burial mound detection using R.

R is an open-source, open-access statistical scripting language which can be used either through the command line or it's GUI, R Studio. It has been used in several scientific domains, and lately it is one of the most commonly used scripting languages in Archaeological Sciences (Schmidt and Marwick 2020, 18). The advantages of scripting languages support the four general principles of reproducible research of Marwick

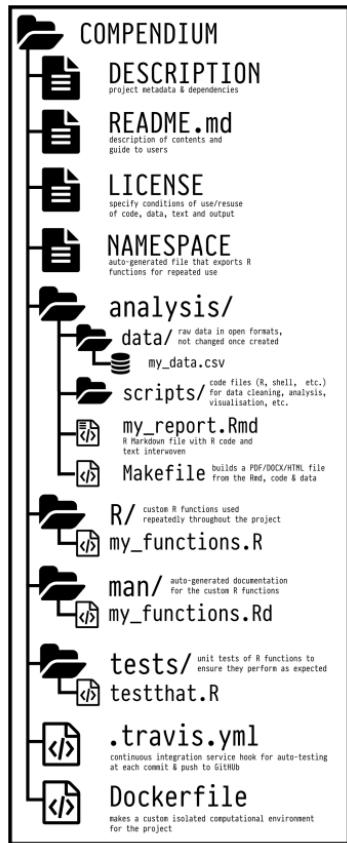


Figure 17: Figure 17. Schematized modus operandi of the Random Forest algorithm.
(#fig:Figure 17)

2017 (see above) which are embodied in an optimal research compendium. The use of scripting languages (can) facilitate the use of Automated Analysis methods in Archaeological Remote Sensing, by offering a clear and logical semantic syntax, delivering a (shared diachronic semantic) ontological consistency throughout the research project and for future use.

Schmidt and Marwick 2020 – after discussing idea- and tool-driven revolutions in science – debate the role of tools in paradigm-shifts in Archaeology. Archaeology is in a paradoxical situation: for one it is mainly an idea-driven discipline, but it has seen multiple tool-driven paradigm shifts (e.g. with Clarke's Analytical Archaeology and the development of Computational Archaeology). Computational Archaeology, Archaeological Remote Sensing and Remote Sensing itself are disciplines where change is mainly tool-driven, as we can see based on Chapter 2, which analysed 31 papers. But tools also ignite ideas, so there is no clear distinction, more a self-induction.

Before going to the next step, the concept of “tool-driven” has to be investigated more thoroughly. Schmidt and Marwick 2020 borrow the concept from Galison 1997, who elaborated on a tool-driven change in particle physics in the twentieth century, meaning tools like digital devices (for instance computers in the case of Archaeology). They see R (or similar open source programming languages) as one projection of a tool-driven revolution, including sharing reproducible and replicable code (Schmidt and Marwick 2020, 19). In my opinion we have to go further - the concept of tool-driven should become more mundane and tangible: automated analysis in Archaeological Remote Sensing has to become tool-driven in order to present a robust scientific practice for Archaeology, which means that reproducible workflows have to be created to do specific tasks, such as feature extraction, segmentation, which can then be developed further and be learned from. Automated analysis in Archaeological Remote Sensing should embrace tool-driven reproducible research to the fullest to make it possible to learn from each other and to build upon the experience of each other to develop ideas further.

One important issue still needs to be addressed. Lately it has been stressed (Strupler 2021) that not all openly published data are comparable: only because the raw data, especially if they are older projects which do not comply yet with the Berlin Declaration on Open Access to Knowledge in the Social Sciences and Humanities, is openly published and the exact methods are known, it does not automatically grant reproducibility (Strupler 2021, 2.). This emphasizes, that not only new data, reproducible workflows and best practices need to be created and published according to guidelines, but also legacy data from on-going, long-term excavations should be revised and curated and updated, to fit the FAIR principles ‘Findability’, ‘Accessibility’, ‘Interoperability’, and ‘Reuse’ (Haas and van Leusen 2020). It is clearly a problematic issue in this case, because often the authors might not be retractable. Thus reproduction of previous studies is a good way to correct, curate and update these datasets and the code and data is perfect for learning and teaching workflows but also to learn how to deal with possible errors (Strupler 2021, 15). It has to be pointed out that it cannot be expected to have complete consistency between a dataset processed in the original study using a GIS platform and then replicated using a scripting language. When using software other than a scripting language (where optimally all steps taken are documented), there are often steps one does not document or forge. An error in a dataset or analysis is a different matter and complicates reproducibility but should not have a negative connotation and should not be seen as a failure in the archaeological scientific community but treated with ‘full disclosure’, as a learning effect on how to make datasets and analysis better. The valid point is raised, that when publishing datasets and analysis there should be a responsible person to turn to (Strupler 2021, 16-17). This is even true when e.g. in a project the data is collected by a different person than the one who will do the analysis. Often crucial information is already missing when doing the analysis in the first place and this also needs to be disclosed clearly in the analysis (metadata in any form suitable for the project) and is not to be left out. In the case of datasets made available in an online repository, the owner of the repository or provider (uploader) of the dataset seems to be the logical person to make responsible if not disclosed otherwise. This can be of course problematic with legacy datasets.

This comes together with the development of best practices and publishing datasets with workflows as mentioned before and fits perfectly in the wider picture of the Reproducible Research Culture (Nakoinz 2021, 63). Reanalysis studies are the perfect means to test reproducibility and the quality of data and metadata (as seen above). Like Quantitative Archaeology, also Automated Archaeological Remote Sensing needs a paradigm shift from a closed and restrictive to a sustainable Open Science.

4 Applied Mound Detection

In certain instances it was decided to use screenshot to be able make points clearer. It was planned to use `mapview::mapview`, but it does not knit nicely when exported as a PDF.

5 Results of the iSEGmound Workflow

In this Master's thesis two data preparation methods (DTm and Multi-Scale Topographic Index) and two segmentation methods (Watershed and Region Growing) were used and compared. The four workflows were tested on the Train DTM and the Train Area to understand the relationship between the size of the area of investigation and the variable settings of the respective algorithms. The most effective workflow was chosen to be applied to the five Areas of Interests: AoI 1, AoI 2, AoI 3, AoI 4 and AoI 5.

First let's have glimpse on the morphometric derivative chosen, the Multi-Scale Topographic Index, on the example of the Train DTM:

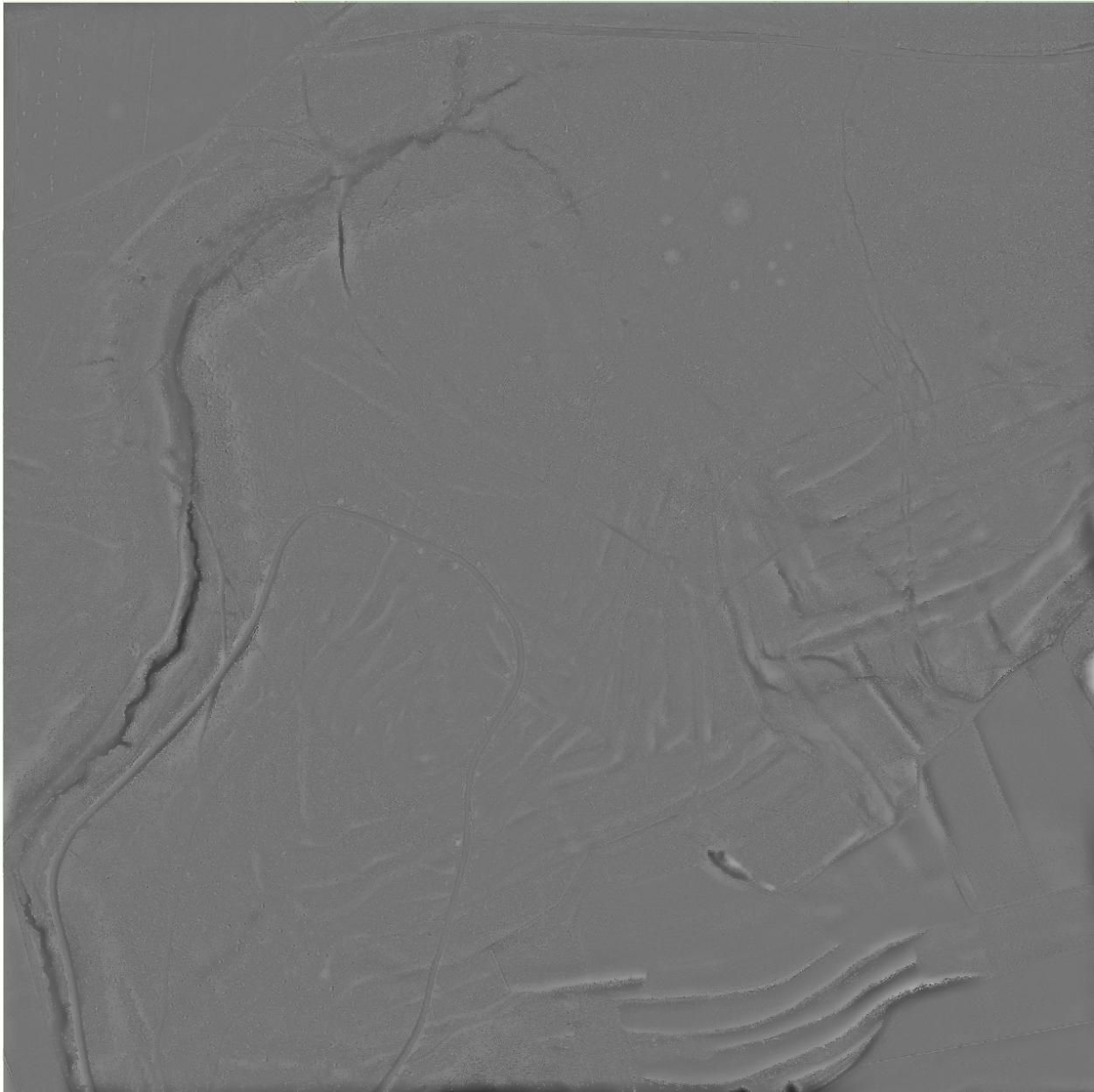


Figure 18: Multi-Scale Topographic Index of the Train DTM.
(#fig:Figure 67)

As a reminder let's see where the burial mound groups Site ID 5 (black) and Site ID 35 (blue) are located in the Training DTM:

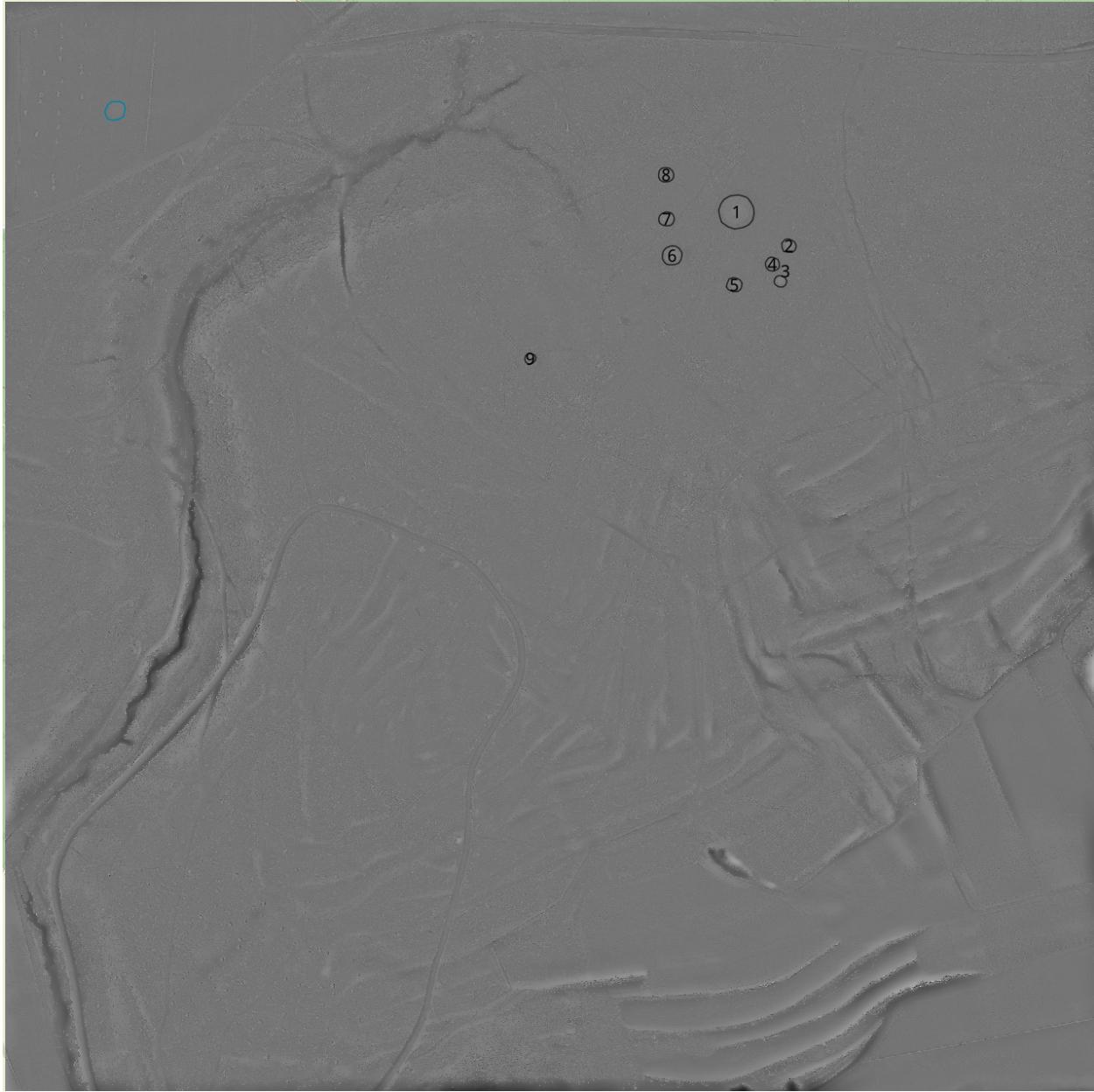


Figure 19: Multi-Scale Topographic Index of the Train DTM with bruijal mound groups Site ID 5 and 35. (#fig:Figure 68)

We know from Dobiat et al. 1994, that Site ID 35 was identified as two mounds. As in Chapter 4 discussed, the mounds visible in Figure 68 were possible to be identified on ground.

5.1 Results of the Training DTM

The workflows applied on the Training DTM are the following: 5a_iSEG05_WS, 5b_iSEG05_mtpi_WS, 5c_iSEG05_RG, and 5d_iSEG05_mtpi_RG.

Let's plot the results of the Training DTM by segmentation. First the Watershed Segmentation based on a DTM (iSEG05_WS, orange segments) and on the SAGA MSTPI (iSEG05_mtpi_WS, lilac segments):

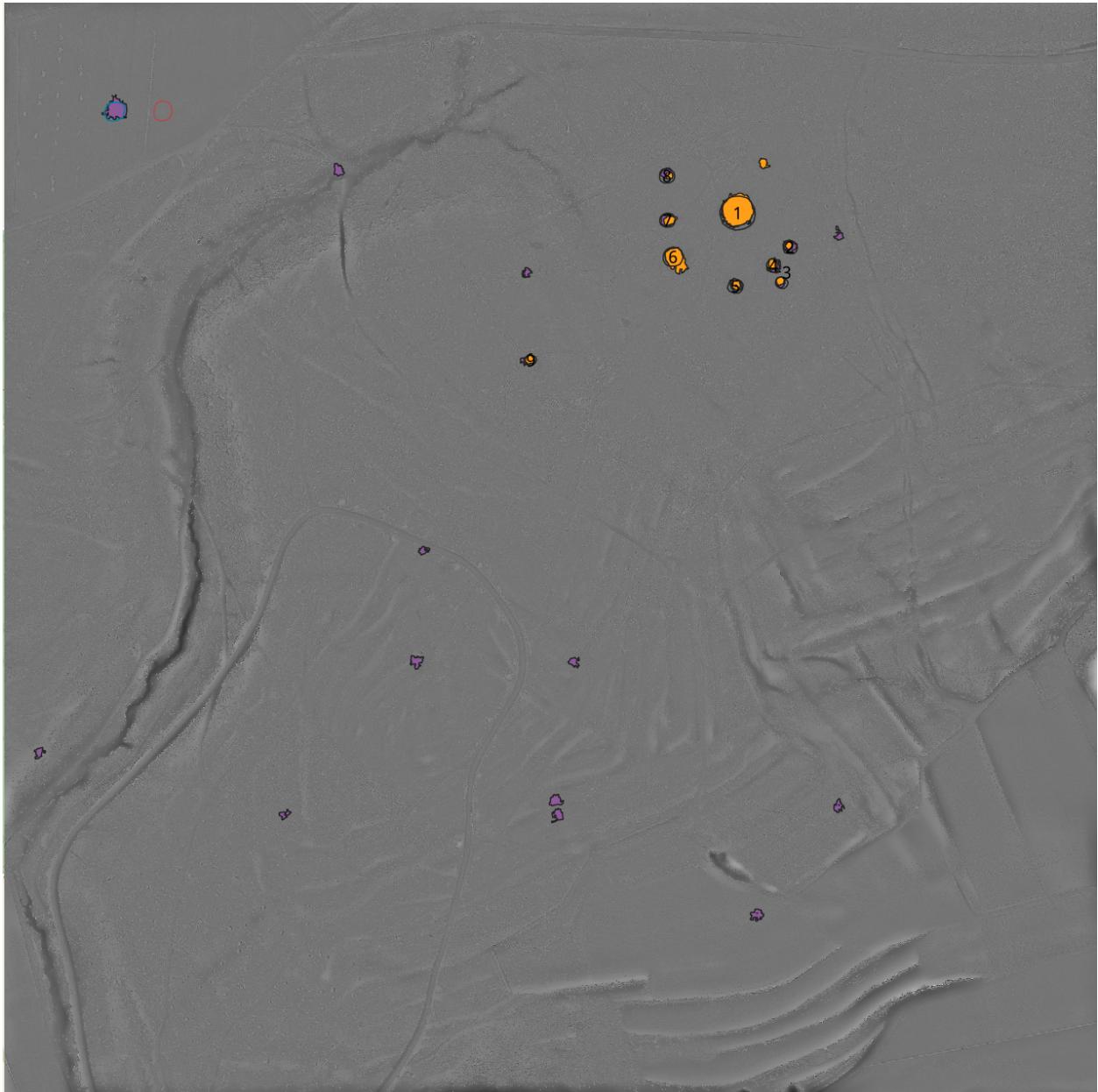


Figure 20: Plotting iSEGWS and iSEG mtpiWS on the SAGA MSTPI.
(#fig:Figure 69)

Now, the Region Growing Segmentation based on a DTM (light blue) and on the SAGA MSTPI:

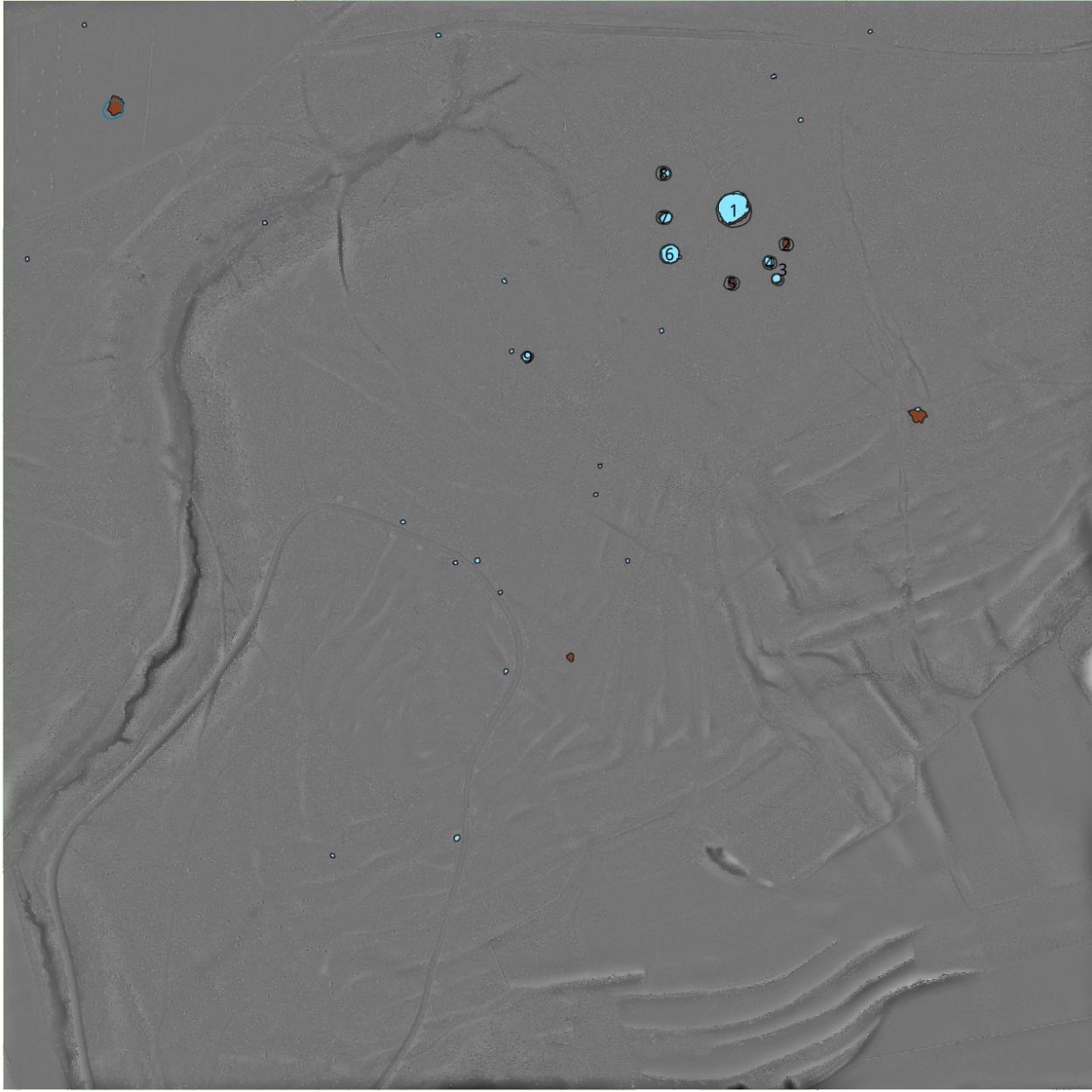


Figure 21: Plotting iSEGRG and iSEGmtpiRG, on the SAGA MSTPI.
(#fig:Figure 70)

From the first results it is clearly visible, that using

5.2 Results of the Training Area The workflows applied on the Training DTM are the following:

6a_iSEG05_WS_ta 6b_iSEG05_mtpi_WS_ta 6c_iSEG05_RG_ta 6d_iSEG05_mtpi_RG_ta

5.3 Choosing the right segmentation

To choose the best working workflow for the Training Area the question posed was: Which segmentation is better? Two different considerations were investigated: the archaeological decision and the statistical decision. From the archaeological point of view the aim is to detect all (locations of) burial mounds.

This can be of course broken down to the question do we want to find the exact shape of the mounds (in the case of the Training DTM and Training Area) or the most important is to detect as much as possible in any shape (e.g. just half or $\frac{3}{4}$ of a mound is detected) but to detect as as possible of them.

6 Discussion

7 Conclusion