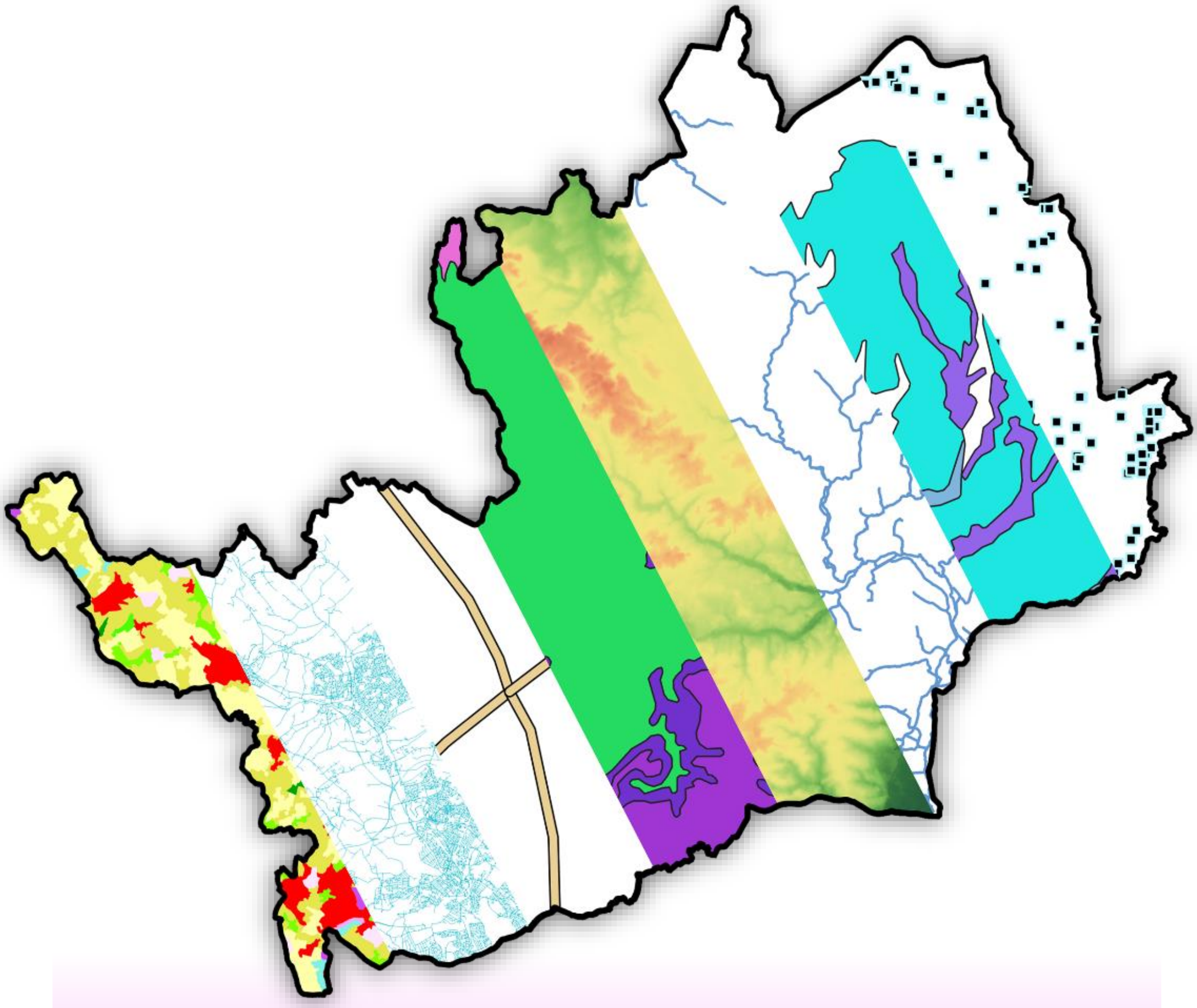
A Space for Predictive Modelling in England's Archaeological Heritage Management System:

A case study of modelling site patterns in Roman Hertfordshire, England



Jennifer Stacey

Figure 1: A collaboration of different data layers that were used to influence and create the Roman Hertfordshire predictive model. From left to right: land-use, modern roads, Roman roads, bedrock geology, digital elevation model, modern rivers, superficial bedrock, archaeological sites (Stacey 2020).

E-mail: jenniferstacey@hotmail.co.uk

A Space for Predictive Modelling in England's Archaeological Heritage Management System:

A case study of modelling site patterns in Roman Hertfordshire, England

Jennifer Stacey

S1829424

MSc Thesis Archaeology (4ARX-0910ARCH)

Supervisor: Dr. K. Lambers

Specialisation: Archaeological Science

University of Leiden,
Faculty of Archaeology

Leiden, 21/08/20

Final draft

Contents

Acknowledgements

I would like to express my warmest gratitude to all those who helped this research come to fruition.

To my supervisor, Dr. Karsten Lambers, for his continued guidance and feedback throughout the writing of my Master's thesis.

To the Archaeological Data Service, Historic England, the Ordnance Survey, British Geological Survey, the European Copernicus Land Monitoring Service, the Office for National Statistics, the North Hertfordshire District Council and Harvard University for providing the data that was used within this research.

To Dr. Milco Wansleeben, for his help and effort in focusing my research ideas.

To the Archaeology Faculty at Leiden University, for the invaluable experiences and the many inspiring people there.

To my family and my friends, for their unending support and encouragement.

Lastly, to Christodoulos Makris, for his love and patience.

1. Introduction

This thesis aims to apply and test the method of archaeological predictive modelling in the context of Roman Hertfordshire, England by using open-access data and applications. Through this application, the potential of integrating predictive models within the current English Archaeological Heritage Management (AHM) system will be discussed. By doing so, the benefits and drawbacks of archaeological predictive modelling can be identified through an applied case-study.

1.1. Brief Introduction to Archaeological Predictive Modelling

The act of archaeological predictive modelling has generally been defined as a set of techniques which are employed to predict “the location of archaeological sites or materials in a region” (Kohler & Parker 1986, 400). This can be done either inductively from “a sample of that region”, or deductively by basing predictions on “fundamental notions concerning human behaviour” (Kohler & Parker 1986, 400). This method has been employed either as a useful tool for archaeological research (Danese *et al.* 2014, 42) or as part of a cultural heritage management strategy as it can create areas of differing “archaeological potential” (Carleton *et al.* 2012, 3371)

Archaeological predictive modelling has been criticised since its evolution for its usage within governmental land management projects in the USA, from the late 1970’s (Kamermans *et al.* 2004, 5). Most criticism addresses the inductive, data-driven approach as it is more prevalent in predictive modelling. The reductionist (Nakoinz 2018, 105) or ecologically deterministic mapping of the historic landscape has also been criticised (Kamermans *et al.* 2004, 6). It can also be argued that predictive modelling only predicts the “relative suitability” of land areas for historic habitation (Verhagen & Whitley 2020, 235), rather than the archaeological reality.

Questioning the underlying theory which shapes a predictive model can be understood through the model's intended purpose, whether that would be for archaeological research or heritage management. These issues found within explanations of human behaviour are perhaps less relevant if the model is intended to be used only to predict the archaeological potential for development purposes, where archaeological potential is often deemed as levels of 'risk'.

1.2. The Motive for the Research

Due to the UK's signing of the Valetta Treaty in 1992, legislative policies were created to protect the archaeological environment from urban developments that are increasing the risks to national heritage (Council of Europe 1992, 4). Within England, the current legislative policies for archaeological heritage are within a single document, 'The National Planning Policy Framework' (NPPF) which was published in March 2012, and updated in 2019. The NPPF superseded earlier legislation that was put in place to implement the Valetta Treaty, such as the Planning Policy Statement 5: 'Planning for the Historic Environment' (PPS5, 1994).

The policies within the current framework express that designated areas of archaeology should be protected, as according to the Valetta Treaty, but leaves much room for interpretation for areas where archaeology likely exists but has not been formally 'designated' (Secretary of State for Housing, Communities and Local Government 2019, 56). The earlier and now out-dated policies of the PPS5 favoured the *in situ* preservation of heritage assets and emphasised the role of the Historic Environment Records (HERs) system in England (Flatman & Perring 2012, 4). The goal of researching and publishing archaeological findings within conservation strategies were also encouraged in the PPS5, whereas the NPPF opts for conservation by any means, preferably at a low cost (Flatman & Perring 2012, 7).

Both the current and past legislative policies regarding heritage management provide much of the motivation to investigate what the benefits may be of using archaeological predictive modelling in the earlier steps of the heritage management system within England. The Netherlands, a nearby neighbour of England, have implemented predictive modelling within their heritage management system on a national scale. Government-backed guidelines require predictive values to be created for an area before a development can be permitted (Willems & Brandt 2004, 28). These values can provide a baseline for deciding which action should be taken to minimise the risks of archaeological disturbance, but also to minimise delays in developments due to the unexpected discovery of archaeological remains.

If standard guidelines are required for the creation and publication of English archaeological predictive models, many of the common criticisms can be addressed. The implementation of these models can provide a less expensive form of additional guidance for both the developer and local authorities responsible for the protection of archaeology.

1.3. Aims and Research Questions

The main aim for this research project is to use the archaeological landscape of Roman Hertfordshire as a case study for investigating the application of archaeological predictive modelling in England for heritage management purposes.

The research involves the collection of accessible, open-source data to inform the predictions, such as the geology, topology, elevation, hydrology and Roman road systems of Hertfordshire, which are collected from various sources. Collection of data also includes the access of known Roman archaeological data within Hertfordshire in order to partially create and test the final predictive model. Finally, the model was tested and discussed in terms of its potential applicability to the Archaeological Heritage Management (AHM) system within England.

The research questions that aim to be answered through the creation and discussion of the Roman Hertfordshire predictive model are the following:

1. Does England have the open-access digital infrastructure[1] to facilitate the creation of well-informed archaeological predictive models?

2. What knowledge can be gained from the creation and output of the Roman Hertfordshire predictive model?

    i. What methodological knowledge about archaeological predictive modelling could be gained from the creation and output of the predictive model?

    ii. What archaeological knowledge about Roman Hertfordshire could be gained from the output of the predictive model?

3. How can the case study of Roman Hertfordshire assess the potential of archaeological predictive modelling within the Archaeological Heritage Management system in England?

1.4. Thesis Outline

Chapter Two provides contextual information on the research areas of Hertfordshire, such as its suitability for the research area, the modern geographical characteristics and a short history of its Roman occupation. The chapter also provides background information on the current Archaeological Heritage Management (AHM) system that is used within England.

[1] The term 'digital infrastructure' is used to refer to the digital data and resources which are available for England and have been granted open-access to the public. This infrastructure can be provided by companies in the UK and the EU, or by the UK government.

Chapter Three introduces the materials that were used to inform and model the Roman Hertfordshire site predictions. The origin of each data layer is given, and each layer explained as to its relation to the environmental or social development of Roman Hertfordshire. More contextual information on historic Hertfordshire is provided through the discussion of the elevation, soil, geology, hydrogeology, river system and Roman road system. The materials also include data on modern Hertfordshire which are also discussed for their relevance, such as protected areas and monuments and modern land-use. The process of 'data cleansing' the known archaeological sites in Roman Hertfordshire is explained, in addition to the sampling and categorising processes that took place.

Chapter Four, explains the predictive factors that were integrated into the Roman Hertfordshire archaeological predictive model, along with the mixture of modelling methods that were employed. The chapter then assesses the applicability and quality of the environmental and archaeological data that was used in the model. A series of suggestions for future improvements that could be made to the selection of environmental and archaeological data are briefly explored.

Chapter Five details the application of the methodology that is explained in Chapter Four, and clearly displays each step of the modelling process. Firstly, the proximity of rivers and Roman roads are evaluated for their potential in predicting Roman sites. Secondly, multi-criteria analysis is conducted on the factors of proximity to water sources and the Roman road network through a weighted procedure. Thirdly, the factors of optimal aspect and slope are integrated into the model through another instance of weighted multi-criteria analysis. Fourthly, site density analysis is used to identify the location of major and minor Roman towns in order to create proximity buffers around each area. The final product is then evaluated by applying the testing sample to the result and by calculating Kvamme's Gain scores with the testing and modelling sample. Applications of the Roman Hertfordshire model are explored through the creation of a developer guide and proximity-based analysis by site types.

Chapter Six provides further discourse on the guidance provided in Chapter Five regarding the application of the Roman Hertfordshire predictive model. Other issues pertaining to the production of archaeological predictive models are also discussed, such as sources of funding in England and the standardisation of their production and publication.

Chapter Seven provides a synopsis of three main research questions which the research aims to answer. It first explores whether England had the open-access digital infrastructure to facilitate the creation of an informed Roman Hertfordshire predictive model, concluding that a sufficient amount of data was available to the public but could have been of higher quality. The chapter then discusses the archaeological and methodological knowledge gained from the creation and final product of the predictive model, overall stating more knowledge was gained methodologically. Finally, the case study of the Roman Hertfordshire predictive model is evaluated in terms of its ability to assess the potential of the method within the AHM system in England. This research question was partially addressed through an explanation of the weaknesses of the method for AHM purposes, specifically within England. However, a proposed 'starting point' is suggested for the method's implementation into the current system, such as applying it to areas with little to no previous archaeological knowledge.

2. Hertfordshire and Archaeological Heritage Management (AHM)

2.1. Characteristics of Hertfordshire

England is divided into forty-eight ceremonial counties, or shires. Thirty-nine of these counties were officially established on the grounds of their cultural, administrative or geographical boundaries sometime in antiquity, and have thus come to be known as historic counties. Hertfordshire is one of these historic counties, located in the south-east of England (fig. 2), and was chosen to be the research area for the model.

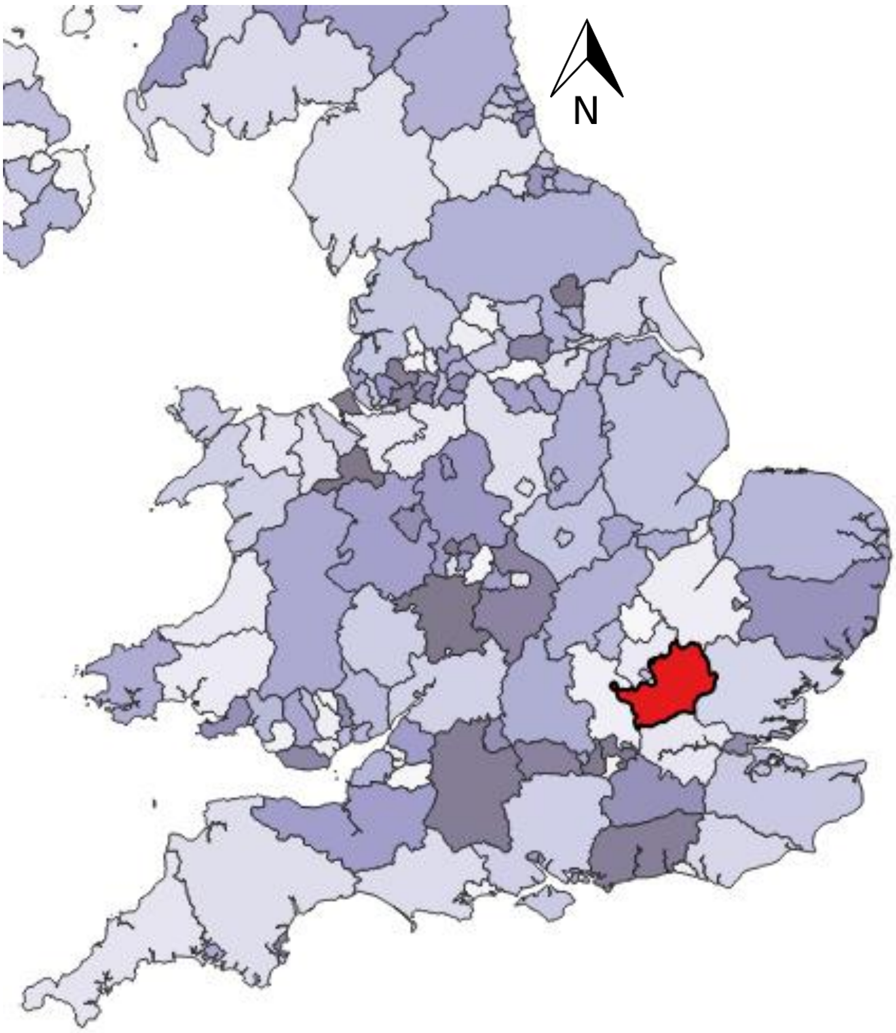


Figure 2: Map of the counties of England, with Hertfordshire highlighted in red. Based upon the 'Counties (April 2019) EN BFC' data source, with the permission of ONS Geography Open Data.

While these administrative divisions did not exist throughout much of history, it was important that the predictive model had a spatial boundary to ensure an acceptable resolution could be achieved in the final product. Therefore, it was decided that the archaeological predictive model would be limited by the modern boundaries of a singular county. It should be stated that these county boundaries, proposed for the boundaries of the research area, do pose theoretical issues to the model as site location was likely to have been influenced by environmental and social factors that lie outside the modern limits.

The number of archaeological data records on the Archaeological Data Service (ADS) was unequally distributed across each county and time period. Therefore, it was my initial task to select a county which was not too large in size, but also had a large amount of archaeological data available among a single archaeological period. The amount of data for each county was determined by the number of search results on the 'ArchSearch' function on the ADS website (www.archaeologydataservice.ac.uk/archsearch). Observing the different search counts led me to consider the southern English county of Hertfordshire, finding that it has a long history of settlement since the Neolithic age. Hertfordshire stood out as having 9263 archaeological data results across all periods, with 1352 of them dating to the Roman period. According to the Office for National Statistics, Hertfordshire ranks 36th of 48 by the size of counties in England at the size of 1,643 km^2 (www.geoportal.statistics.gov.uk). Therefore, the relatively small size of Hertfordshire, along with a large amount of Roman archaeological knowledge stored by the ADS, proved the county was a good candidate for my predictive model case study.

Hertfordshire borders five counties; Cambridgeshire and Bedfordshire to the north, Essex to the east, Greater London to the south and Buckinghamshire to the west (fig. 3).

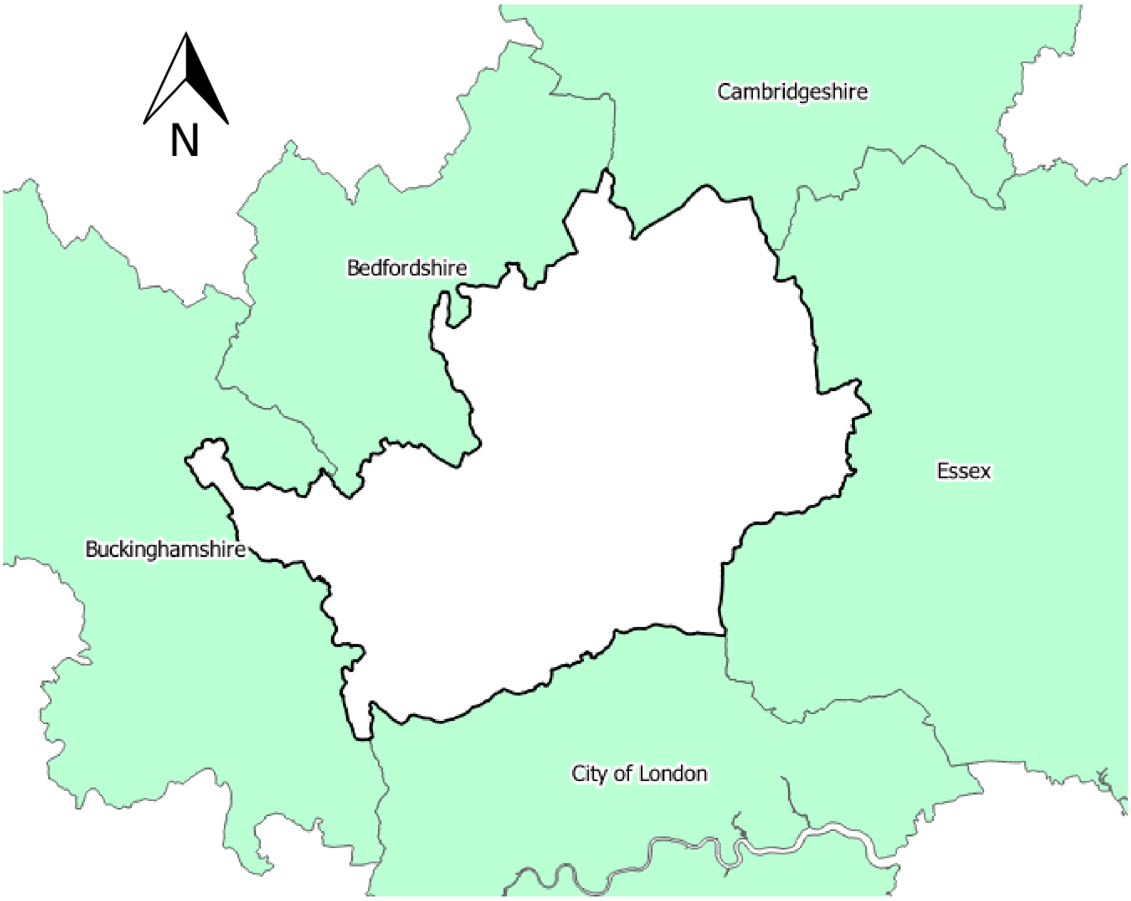


Figure 3: The five counties which border Hertfordshire. Based upon the 'Counties (April 2019) EN BFC' data source, with the permission of ONS Geography Open Data.

The county of Hertfordshire consists of ten districts, namely North Hertfordshire, Stevenage, East Hertfordshire, Welwyn Hatfield, Broxbourne, St. Albans, Hertsmere, Watford, Three Rivers and Dacorum (appendix 1).

While Hertfordshire is still considered a rural county, increasing population and household growth demands lead to ever expanding urbanisation of the landscape. This is especially accelerated by its bordering position next to Greater London, whom has been expanding over time. More information on the geological and hydrogeological characteristics of Hertfordshire is provided in the 'Materials' chapter which documents the different geological formations and deposits that make up Hertfordshire.

2.2. Roman Occupation of Hertfordshire

Hertfordshire has a long history of occupation dating to the Neolithic period, which was characterised by the creation of long barrows as ritual monuments (Tereszczuk 2004, 10). This early activity was concentrated around the "proto-Thames" and the river valleys. During the Bronze Age, "significant areas of woodlands" were cleared by the inhabitants, continuing through the Iron Age (Dacorum Borough Council 2004, 6).

By the late 40's AD, the Romans "almost held all of south-west Britain", but the conquest of south-east Britain likely took much longer (Menard 2011, 44), placing the conquest of the area that is now Hertfordshire between the years of 43-84 AD (Menard 2011, 46). This conquest of the land by the Romans brought major changes to the landscape of England, and what is now the area of Hertfordshire. Between the years of 50-60 AD, the revolt of the Iceni, a tribe of British Celts, resulted in the destruction of the town of Verulamium (Menard 2011, 46), which at one point was named the third largest town in Roman Britannia (Lockyear & Shlasko 2017, 17). These tumultuous periods within the consolidation of Roman rule are closely tied to the development of the Roman road system (Menard 2011, 47), in addition to other forms of landscape modifications.

2.2.1. The developing Roman landscape

Large-scale road networks were brought to Britain for the first time in its history, constructing at least four major road networks that passed through the area of Hertfordshire. These roads connected the existing *municipium* at Verulamium (St Albans) (Historic England 2018, 1; Rogers 2013, 4) as well as the Roman towns of Welwyn, Braughing and Ware (Dacorum Borough Council 2004, 6) to the larger landscape. Landscaping for recreational purposes was introduced, probably growing "an avenue of trees and shrubs" (Dacorum Borough Council 2004, 6) lending a hand in 'Romanizing' the environment.

The first evidences of drainage systems were brought to Britain upon Roman conquest (Brown 1997, 269) along with other forms of water management (Historic England 2018, 4). Systems of Roman water management were identified at the Roman town of Braughing in Hertfordshire (Brown 1997, 226). Excavations of the Roman Gate at St Albans, Hertfordshire looked at the site of Verulamium. The investigation identified clear evidence of a man-made redirection of the river Ver in order to control flooding. The creation of this new water channel around the Roman town would have required cutting into solid rock and creating a levee (Rogers 2013, 63). Remarks have been make on the kinds of labour and expertise needed to enact this kind of environmental change (Rogers 2013, 119). Wooden water pipes were also found in Verulamium, suggesting that water access was facilitated not only by wells but "brought into the settlement by an aqueduct" (Rogers 2013, 133). This manipulation of the landscape and the creation of new forms of water access probably greatly affected the choice of site location within Roman Hertfordshire.

2.2.2. Roman sites

Through the excavation work that has taken place in the last century, many Roman sites have been identified within Hertfordshire. Due to increasing modern development, there has been a sharp increase in archaeological work undertaken in Britain (Holbrook 2015, 1), leading to the uncovering of Britain's Roman past. Within this research, many of these finds are categorised by their function, pertaining to their involvement with either settlement sites, economic sites (agricultural or industrial), ritual sites (funerary or temples) or military sites.

Roman settlements within Hertfordshire can be largely classified as rural settlements (Historic England 2018, 3; Taylor 2013, 173), with both minor and major towns among them. The site of Verulamium (St. Albans) is the largest Roman settlement within the area of Hertfordshire (fig. 4), and is currently "the only certain example in England" of a Roman *municipium* (Historic England 2018, 1), a status possibly granted as an upgrade from a *civitas*-capital (Rogers 2013, 4). Other minor settlements included Ware, Welwyn and Baldock (fig. 4). Within archaeological research, there has been an emphasis on major towns in Roman Britain (Holbrook 2015, 1), perhaps leading to unequal surveying and discovery. This is perhaps through research bias or simple issues of visibility and ease of discovery. However, through the use of predictive modelling, perhaps a more equal, overview can be gained from the landscape as to their potential for holding Roman archaeology.

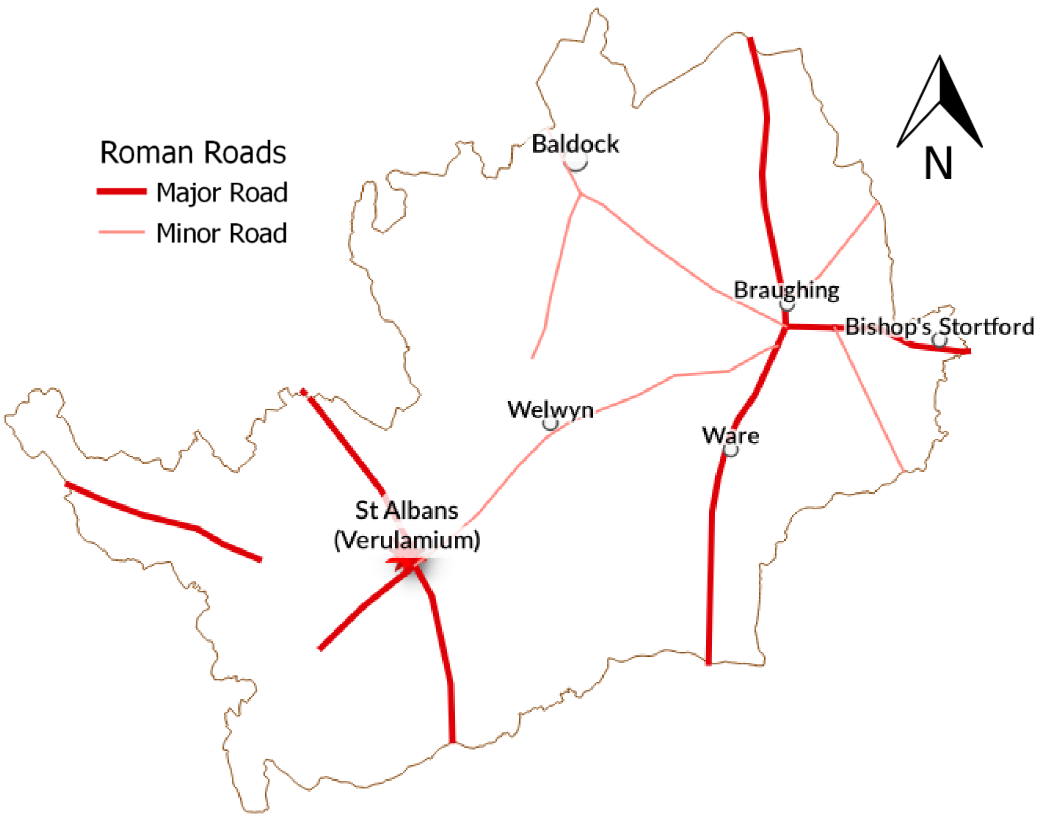


Figure 4: Roman towns within Hertfordshire, connected by a series of major and minor roads. Most of the Roman settlements in Hertfordshire are referred to by their modern name as the Roman name is unknown.

Roman economic sites within Hertfordshire had an emphasis on agricultural and craft – both industrial and domestic – like much of Roman Britain (Taylor 2013, 173). Villa structures were often associated with rural landscapes where agriculture would occur within the villa's estate (Historic England 2018, 2; Taylor 2013, 173). Other types of sites in the Roman countryside also included mining complexes (Historic England 2018, 2) in which the chalk plateau which covers much of Hertfordshire might have influenced this type of site location.

Ritual sites compromised another occurrence in the Roman countryside, specifically temple complexes. Roman ritual-related hoards have been found in the village of Ashwell, four miles north of Baldock. The collection of concealed precious metal objects have been interpreted in multiple ways since its discovery. However, its burial place was "intimately linked to a ritual site" and made it likely the hoard was religious in nature (Jackson & Burleigh 2018, 29). It has been theorised that its existence is evidence of Romano-British pagan shrines (Jackson & Burleigh 2018, 30). Other ritual-related hoards of coins have also been found within the town of Baldock (Phillips *et al.* 2009, 113). A Romano-Celtic temple was identified within Verulamium, constructed in the early Flavian period (69-79 AD) (Fulford 2015, 63), which is likely one of many in the area of Hertfordshire. Burials and cemeteries were normally located on "the approach roads" due to Roman legal requirements that graves are made outside of settlements (Historic England 2018, 8), however this was likely to have varied based on the settlement size and centrality. New-born children were also a known exception to this rule, often found buried within settlements (Historic England 2018, 8).

Military sites are perhaps harder to define as some form of defence could have appeared in a settlement only after a certain period, or in ways that are not visible archaeologically (Historic England 2018, 8). It is likely that military sites were a hybrid-type of site, combining settlement or economic aspects with defensive elements.

2.3. Archaeological Heritage Management (AHM) Practices

2.3.1. The Valletta Treaty effect

The ‘Convention for the Protection of the Archaeological Heritage of Europe’, also known as the Valletta Treaty (1992), is an international convention that has been adopted by forty-five members of the Council of Europe (www.coe.int). It was made to replace the older ‘European Convention on the Protection of the Archaeological Heritage’, or the London Convention (1969). The London Convention dealt with the protection of archaeological heritage through the creation of “reserve zones” and focused on prohibiting “illicit excavations” (www.coe.int) during a time which vandalising archaeological sites was perhaps more commonplace.

Over 20 years had passed since the London Convention in 1969, and the issue of increasing urbanisation and the demand for large-scale development projects created a situation where archaeology was no longer at risk by clandestine excavation, but at risk of destruction by major public works (Council of Europe 1992, 1). The Valletta Treaty of 1992 sought to address a shift of priorities in regards to archaeological protection. The ways in which the treaty advocated for this new protection is central to understanding the Archaeological Heritage Management (AHM) practices that occur within England as well as other countries within Europe.

Legislative policies were required to be made in every country who signed the treaty, requiring a legal system that sought protection of archaeological heritage (Council of Europe 1992, 4). This meant that any operation which intended to disturb the soils below cannot be allowed unless it was previously cleared by the relevant authorities. These positions of authority were created with varying systems in each signing country. In the case of England’s response to the treaty, the local authorities of each county were mostly responsible for determining planning permissions (Chartered Institute for Archaeologists 2017, 5; www.archaeologists.net/find/clientguide).

Another central change the treaty had created in AHM was within Article 6, stating that the responsibility of funding the necessitated archaeological work was to be placed upon the development companies (Council of Europe 1992, 6). This created a situation where developers could be granted planning permissions in an area, only to encounter unexpected discoveries during the work itself. This creates the problem where developers are forced to pay for the required archaeological excavations or to abandon the project. Without an efficient AHM system, this situation is likely to happen often. For the developer, this can lead to severe project delays and a loss of profit. For the archaeologist, the situation can also cause problems. Commercial development companies in Britain have come to provide 'lump sum' contracts to archaeologists to do the unexpected work (Heaton 2014, 246), thereby leading to underpaying for extensive excavations. It is thought that with a better risk management toolkit, better estimates on the cost of value of work can be made and would help avoid this problem (Heaton 2014, 246).

Other forms of legislation also exist within England, as well as the rest of the United Kingdom (Wales, Scotland and Northern Ireland) that aim to protect archaeology. One of the central legislations is 'The Ancient Monuments and Archaeological Areas Act' of 1979, which provides two main forms of protection for archaeology by prohibiting the disturbance of scheduled monuments and 'areas of archaeological importance' (Benetti & Brogiolo 2018, 181; www.legislation.gov.uk/ukpga/1979). Monuments are selected to be 'scheduled' on the basis of their "national importance" which is assessed by its period, rarity, condition, and vulnerability (historicengland.org.uk). Areas deemed to be of archaeological importance are decided by the local authority by their value. Any disturbance to a scheduled monument or the soil of an archaeological area will result in a criminal offence to the parties involved.

2.3.2. Planning permissions and the assessment process

The current system of planning permissions within England comes from the National Planning Policy Framework. The subsequent assessment process which follows the planning application in England is facilitated by foundations such as Historic England (formerly English Heritage) and the Chartered Institute for Archaeologists (CIFA) who have provided documentation on the standards of assessment. Across the rest of the United Kingdom, alternative planning policies are used, such as the 'Planning Policy Wales 10' (December 2018), the 'Strategic Planning Policy Statement for Northern Ireland' (September 2015) and the 'Scottish Planning Policy' (June 2014).

The National Planning Policy Framework (NPPF) was published in 2012 and has since been updated in 2019. The implementation of the NPPF was created to replace a wide range of planning policy statements and guidelines within a single document, such as the 'Planning Policy Statement 5: Planning for the Historic Environment' (PPS, 1994) and the 'Planning Policy Guidance Note 16: Archaeology and Planning' (PPG, 1990) (Flatman & Perring 2012, 4; www.designingbuildings.co.uk). The sentiment of the NPPF has been said to promote the agenda of 'localism', by aiming to "put power back into the hands of local people" (Flatman & Perring 2012, 5).

The NPPF assigns Chapter 16 to discussing the conservation of the historic environment. The policy states that conservation of heritage assets should be equal to their significance – significance which is deemed by the local planning authorities (Secretary of State for Housing, Communities and Local Government 2019, 55). If a proposed development will lead to substantial harm to a "designated heritage asset", local authorities have the obligation to refuse planning consent (Secretary of State for Housing, Communities and Local Government 2019, 56).

However, the document has been debated since its implementation in regards to its impact on heritage management. Direct comparisons have been made between the NPPF and the PPS. For example, within the PPS the actions promoted should be ‘in favour of conservation of heritage assets’, while in the NPPF the statement has been “infamously” changed to be ‘in the favour of sustainable development’ (Flatman & Perring 2012, 6). The importance of ‘designated’ heritage assets being protected is also repeated, leaving guidance on undesignated heritage assets unclear as to how to proceed. The likely outcome will encourage the practice of mostly producing ‘Desk Based Assessments’ for undesignated sites.

Planning permissions are granted by the local authority on the grounds of gathered evidence regarding the impacted area’s historic value (English Heritage 2015, 2). Sources of evidence are to be found on The Historic Environment Record (HER), the National Heritage List for England (www.historicengland.org.uk) or on the Heritage Gateway (www.heritagegateway.org.uk). Local voices are also used in planning permissions, using local history groups and civic societies for additional information on the area’s potential or value (English Heritage 2015, 2). However, this system has led to several English counties basing their heritage management on the proximity of known archaeological data (Wilcox 2014, 341) which can be vulnerable to various biases (Van Leusen 2002, 76; Verhagen *et al.* 2007, 203; Verhagen & Whitley 2012, 85). It is officially advised that only in cases where existing records do not provide enough evidence that an “appropriate archaeological assessment” method should be used (English Heritage 2015, 3).

2.3.3. Predictive modelling and the English AHM system

Research in 2012 on the region of East Anglia (Norfolk, Suffolk and Cambridgeshire) revealed that on average the current system of AHM actually discovers some kinds of archaeological remains in two thirds of development-related investigations (Wilcox 2012, 355). Little has changed in the AHM system since 2012 which begs the question, how effective is this system in the rest of

England? Another point to note would be the expenses which are used in the process of archaeological investigations, with research suggesting it "costs the tax payer and developers" around £1 million per year, per county (Wilcox 2012, 355).

The current system of AHM in England has the potential to put both archaeologists and local authorities in a difficult situation where they are unable to assess the potential damage without a sufficient overview of the archaeological situation. It is for this reason that the potential of using archaeological predictive modelling within the existing system should be measured. This potential is attempted within this thesis, by using the temporal limits of Roman Hertfordshire as a case study. The currently accepted method of predictive modelling is prone to many weaknesses, but if it is harnessed with the appropriate level of theory, testing and standardisation then an AHM system has the potential to become more effective, efficient and streamlined through its use.

Heritage management systems that involve predictive modelling have already been widely implemented in the USA, Canada and the Netherlands, and have been implemented to a lesser degree in Germany, the Czech Republic and Australia (Verhagen & Whitley 2012, 53). Some stated positives to the use of these models include "well-informed and transparent decision-making" (Lauwerier *et al.* 2018), the "cost-saving benefits" (Verhagen & Whitley 2012, 50), increasing of "the yield of archaeological inventories" (Verbruggen 2009, 28) and can avoid the biases of known archaeological observations through deductive modelling techniques (Verhagen *et al.* 2007, 203).

However, archaeological predictive modelling is still a highly contested method due to the weaknesses of its application. Attempts to reconstruct pre-modern landscapes with modern landscape layers are difficult when avoiding implicit biases (Kempf 2019, 126). The simplifications that are sometimes made when modelling have been called "reductionist and pragmatic" (Nakoinz 2018, 105) as well as ecologically deterministic for not also modelling social factors in site locations (Kamermans *et al.* 2004, 6). The production of these predictive models lacks a common standard (Wilcox 2014, 345) to maintain quality, and sometimes lacks the mechanism to test the accuracy or reliability of the model (Kamermans *et al.* 2004, 5).

The Netherlands offers a particularly interesting case in the implementation of predictive modelling within AHM through their creation and, to a certain extent, implementation[2] of a national archaeological predictive model, the 'Indicative Map of Archaeological Values' (IKAW). The IKAW was initially produced in by The State Service for the Archaeological Heritage (ROB) with the expressed purpose to guide planning policies (Kamermans *et al.* 2004, 11). However, critical issues have been found with the national map, such as the lack of information given on the density, age or type of sites found (Van Leusen 2009, 52) and therefore the result may be seen as reductionist. The data sources which were used to create the map have also been heavily criticised for being too "ecologically deterministic" (Kamermans *et al.* 2004, 15), and could be biased towards the modern landscape (Kamermans *et al.* 2004, 12). In addition to this, only parts of the map have been tested (Kamermans *et al.* 2004, 13) which is unfortunately often the case with older archaeological predictive models (Wilcox 2014, 344; Verhagen & Whitley 2012, 56).

A current manual on the use of the third edition of the IKAW was published by the *Rijksdienst voor het Cultureel Erfgoed* (Cultural Heritage Agency of the Netherlands) (www.cultureelerfgoed.nl) in May of 2009, including newer guidelines of its place within Dutch archaeological heritage management. The document advises the predictive model should be used as a global insight during the early stages of planning (Deeben *et al.* 2009, 4) along with other forms of archaeological information, and used later on in the planning process as a means

[2] In 2004, an assessment found that three out of twelve Dutch provinces did not use the national predictive model (IKAW) to coordinate their policies (Kamermans *et al.* 2004, 17). Since 2009, the advised use of the IKAW is to guide planning permissions, when used in conjunction with the Archaeological Monuments Map (AMK) (Deeben *et al.* 2009, 4).

of determining the scope of archaeological research required (Deeben *et al.* 2009, 5). Perhaps this is an ideal use of a national predictive model, as it provides an overview for planning authorities to consider the archaeological risk of a development but also demands further investigation into the specific area concerned.

Ultimately, a successful archaeological heritage management system seeks to document and protect the known and unknown archaeology within a local authority. At face-value, archaeological predictive modelling can be seen as a useful tool to reach this goal, however the application of this tool within the setting of England is very novel and unheard of. The reasons for this rejection were astutely summarised in Wheatley's 2004 article. Wheatley states that the earlier stages of inductive predictive modelling presented much theoretical and methodological issues and problematic biases which were very opposed within the UK (Wheatley 2004). These associations with the methodology continued regardless of the later theory-driven phases of archaeological predictive modelling that began to include social factors (Kamermans *et al.* 2004, 5) and model testing methods (Verhagen & Whitley 2012, 83).

Perhaps a reconsideration of the methodology of archaeological predictive modelling is due within England, as well as a reassessment of its potential benefits to an AHM system. Through the improvement of predictive modelling by addressing its documented weaknesses, there may be merit in the benefits the method can provide the AHM system within England.

3. Materials

The material collection for this project consisted of collecting relevant open-access data which could be used to create the archaeological predictive model of Roman Hertfordshire. This included data on various environmental and social factors which could have influenced Roman site patterns, in addition to data of known Roman archaeological finds within the county. Data came from various sources and in various forms, both of which will be discussed further.

This data was used to create ‘shapefile’ layers within the open-source desktop programme, QGIS (Quantum Geographical Informational System) version 3.6.0, with GRASS (Geographic Resources Analysis Support System) version 7.6.0. QGIS (3.6.0) was chosen to create the archaeological predictive model due to prior experience with the software and its affordability as open-access. QGIS allows the overlaying of both vector and raster layers which was important when using both an elevation model, which is raster-based, simultaneously with various vector-based map layers. The programme also offers the implementation of open-source plug-ins from the internet which allows for the use of specialised tools along with the software’s ‘native’ analysis capabilities. This feature was used with the implementation of the ‘Point Sampling Tool’ plug-in, created by Borys Jurgiel (www.plugins.qgis.org). This tool enabled the collection of raster values at specified sampling points. Besides QGIS, a spreadsheet application (*Microsoft Excel*) and database management system (*Microsoft Access*) were used for the data cleansing process as well as to create tables, graphs and conduct frequency counts.

The materials which were used to make the predictive model included the collection of open-access data sources and the production of a total of thirteen QGIS layers (tab. 1). However, not all of the layers that were made were used explicitly to produce the predictive model. Some layers were used to add modern contextual information (modern roads, modern land-use and protected areas, archaeological areas) about preservation or to improve navigation. In addition to this, some layers were created to provide contextual environmental information (lower bedrock, superficial bedrock, soil texture and hydrogeology) to the area of Hertfordshire.

Table 1: The layers and data sources used to create and inform the Roman Hertfordshire archaeological predictive model, alphabetically ordered. The format, uses and reclass column explain how and why the layers were used.

Layer	*Source*	*Format*	*Use*	*Reclass*
Archaeological Roman Sites	**Archaeological Data Service (ADS)** Query of Roman Hertfordshire on '*ArchSearch*'	CSV, Point	**L**	
Bedrock Geology	**British Geological Survey (BGS)** *"BGS Geology 625k – Bedrock"*	GeoPackage, Polygon	**P/O**	
Digital Elevation Model (DEM)	**European Copernicus Land Monitoring Service (ECLMS)** *"European Digital Elevation Model (EU-DEM) v1.1"*	TIFF, Raster	**L**	
Hertfordshire Boundaries	**Open Geography Portal (Office for National Statistics)** *"Counties (April 2019) EN BFC"* *"Local Authority Districts (December 2019) UK BFE"*	Shapefile, Polygon	**L**	
Hydrogeology	**British Geological Survey (BGS)** *"BGS hydrogeology 625k"*	GeoPackage, Polygon	**L, P/O**	
Modern Land-Use	**European Copernicus Land Monitoring Service (ECLMS)** *"CORINE Land Cover (CLC 2018)"*	GeoPackage, Polygon	**P/O**	
Modern Roads	**Ordnance Survey (OS)** *"OS Open Roads"*	Shapefile, Line	**P/O**	
Archaeological Areas	**Data.gov.uk (North Hertfordshire District Council)** *"Archaeological Areas"* (June 2014)	WMS, Raster	**P/O**	
Rivers	**Ordnance Survey (OS)** *"OS Open Rivers"*	Shapefile, Line	**L**	
Roman Roads	**Harvard University (McCormick *et al.* 2013)** *"Roman Road System (Version 2008)"*	Shapefile, Line	**L**	
Scheduled Monuments	**Historic England** *"Scheduled Monuments"*	Shapefile, Polygon	**P/O**	
Soil Texture	**British Geological Survey (BGS)** *"Soil Parent Material Model"*	GeoPackage, Polygon	**L, P/O**	
Superficial Geology	**British Geological Survey (BGS)** *"BGS Geology 625k – Superficial"*	GeoPackage, Polygon	**L**	

L = relevance in the choice of a location in antiquity
P/O = relevance in the preservation and observability in the present

Layers will be discussed in the context of their relevance in the choice of a location in antiquity, or their relevance for the chance of a site being preserved and observable in the present.

The layer of known archaeological sites within Hertfordshire underwent a considerable amount of data cleansing which will be discussed briefly within the section on the use of ADS data (www.archaeologydataservice.ac.uk). A number of layers were also reclassified in order to make the information more relevant to archaeological contexts, these included archaeological sites, hydrogeology (groundwater), soil textures, rivers and modern land-use.

3.1. Elevation and Derived Layers

The Digital Elevation Model (DEM) used in the predictive model was provided by the European Copernicus Land Monitoring Service (ECLMS). It is a full-coverage raster with a resolution of 25 meters. It originates from a mixture of SRTM (Shuttle Radar Topography Mission) and ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) data which were used together by a "weighted averaging approach" (www.land.copernicus.eu) thus giving it better coverage and accuracy.

Through this elevation model it can be seen in Figure 5 that the lowest parts of Hertfordshire were created by erosion from the waterways which cut into the higher elevated hill areas. Numerous river valleys then drain off incoming precipitation which feed the Thames river catchment area in the lowest elevated south-east of the area. The highest elevated areas in the north of Hertfordshire were deposited in the Quaternary by glacial meltwaters, boulder clay and glacial drift deposits (Tereszczuk 2004, 7). While these quaternary deposits were likely present in the area of Hertfordshire during the Roman era, it should be noted that a modern DEM can only reflect the modern landscape. Through processes of erosion, or alluvial and colluvial deposition near waterways or hillslopes, the elevation within Hertfordshire has gradually changed. Within the Roman age, Hertfordshire was likely more homogenous than as we see it today, with shallower slopes and river banks.

Regardless, elevation data provides derived information such as the varying degrees of slope across a landscape, the degree or direction of aspect and the appearance of hillshade.

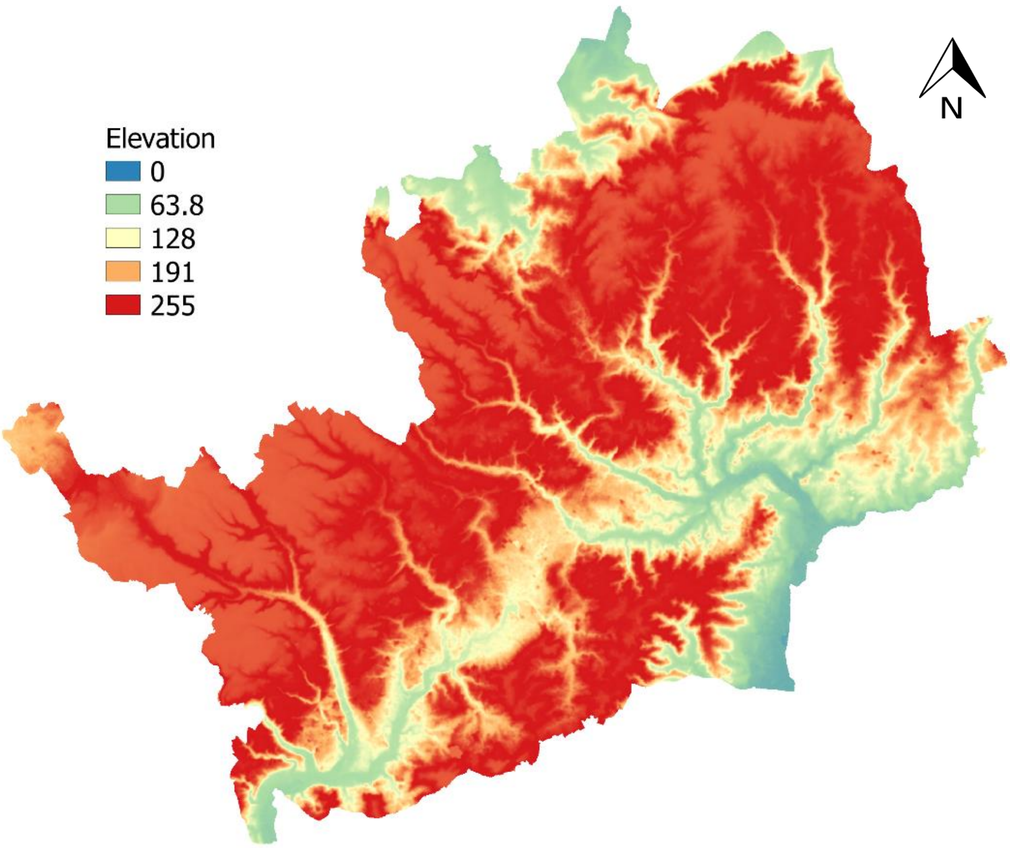


Figure 5: Digital elevation model of Hertfordshire. Based upon the 'EU-DEM v1.1', with the permission of the Copernicus Land Monitoring Service.

3.1.1. Hillshade

The derived layer of hillshade provides a visual overlay of the terrain. In Figure 6, the hillshade layer is placed over the top of the elevation layer, giving a shaded relief effect. While it cannot help predict Roman site locations, the hillshade layer can be used with other layers to provide the same shaded effect.

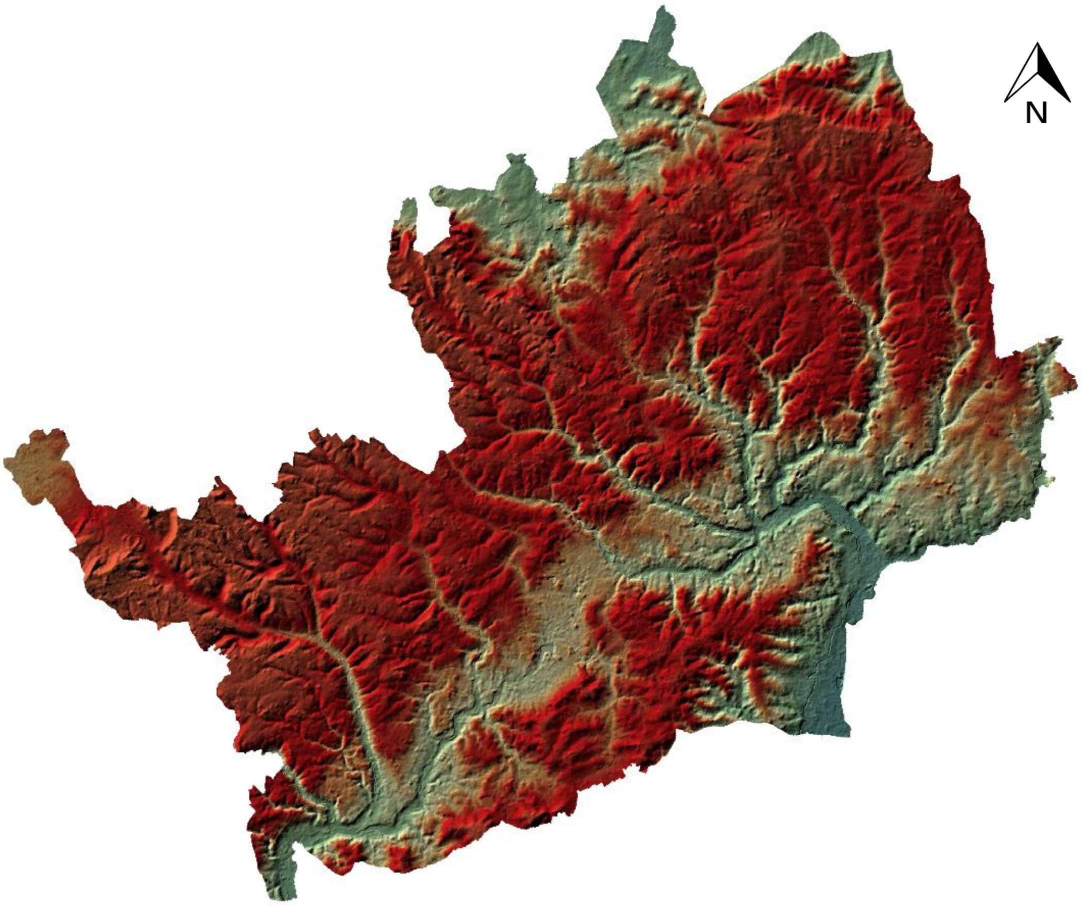


Figure 6: Derived hillshade texture and digital elevation model of Hertfordshire. Based upon the 'EU-DEM v1.1', with the permission of the Copernicus Land Monitoring Service.

3.1.2. Slope

The slope layer that was derived from the elevation data shows where areas of steep and shallow slopes occur by calculating the gradual or sudden change in elevation (fig. 7). Within Hertfordshire, the steepest slopes occur around the eroded areas along the river banks, while in the highest elevated areas of the hills the slope degree remains shallow to none.

The degree of slope can impact a landscape in various ways. Animal husbandry or cultivating crops on steep slopes can be difficult (Wilcox 2014, 341) as the slope may lead to decreased water retention in the soil as the forces of gravity cause it to flow downwards. In addition to this, building structures on very steep ground leads to the need to create foundations, and often impacts the layout of a settlement drastically. Due to this, the assumption can be made that in most cases in antiquity areas with a lower slope degree may have been sought after.

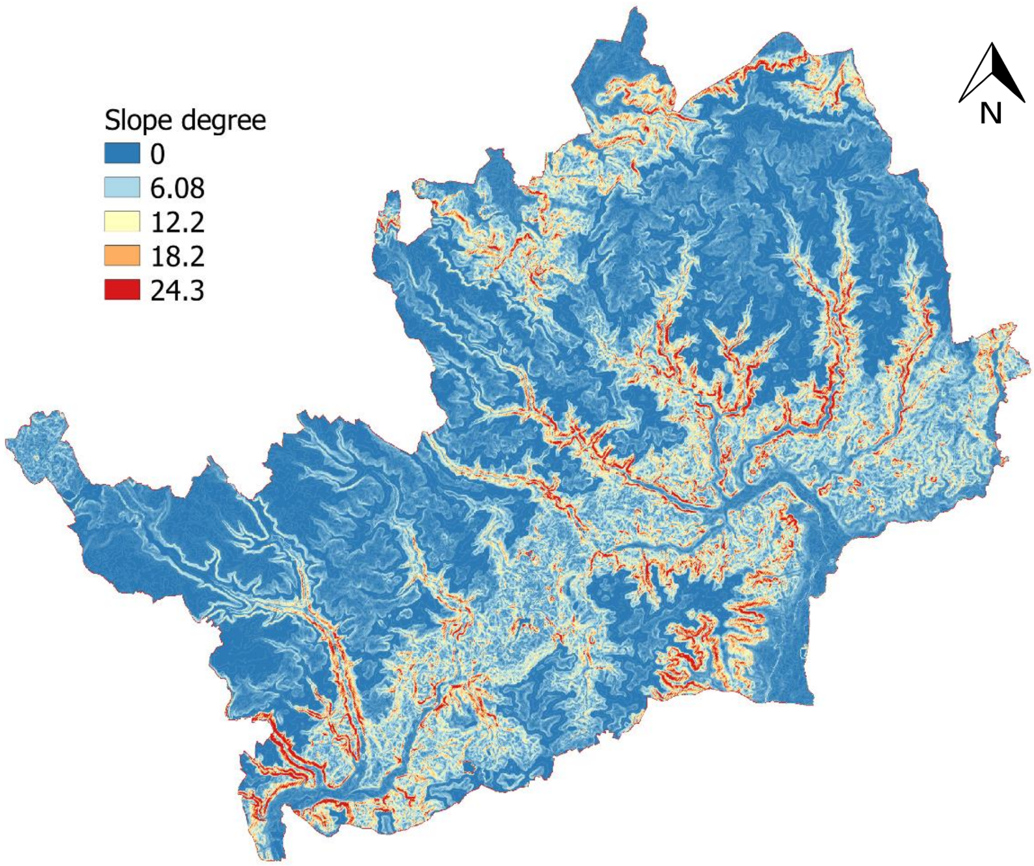


Figure 7: Derived slope model of Hertfordshire. Based upon the 'EU-DEM v1.1', with the permission of the Copernicus Land Monitoring Service.

3.1.3. Aspect

The derived layer of aspect indicates which areas of a hilly landscape can receive the most or the least amount of solar radiation (fig. 8). As England is within the Northern Hemisphere, the degree of aspect which would receive the most sun and the least shaded time would be anything including the southern facing degrees. However, the aspect degree does not matter in areas where there is very low slope, and therefore both the slope and aspect layer should be used in conjunction with each other.

Within Roman Britain, settlements are often associated with the rural landscape, in which even towns are used as hubs for farming (Historic England 2018, 8). Therefore, there are many benefits for the building of sites within these southern facing areas as the sun is depended on for the most basic and complex systems of agriculture.

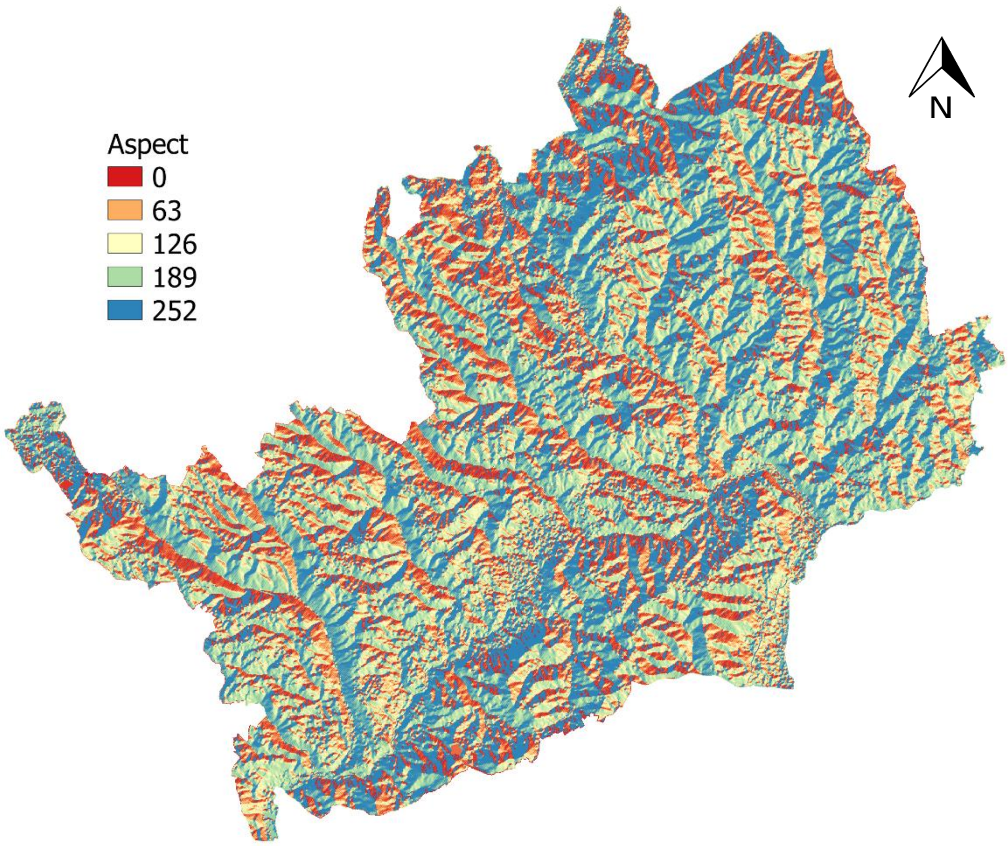


Figure 8: Derived aspect model of Hertfordshire. Based upon the 'EU-DEM v1.1', with the permission of the Copernicus Land Monitoring Service.

3.2. Environmental Layers

3.2.1. Soil textures

The layer of soil textures can be relevant for the prediction of undiscovered archaeological sites, as well affect the preservation and observation of archaeology. It was provided by the 'Soil Parent Material Model', created by the British Geological Survey (BGS) group (www.bgs.ac.uk). The freely available version of the model was only available with a resolution of 1000 meters, and created a pixelated image when used for the county of Hertfordshire. Modern UK digital soil maps often only display basic soil properties which are not influenced by fertilisers or drainage systems, and therefore are able to recreate the properties that are similar to historic soils (Wilcox 2012, 355).

In antiquity, environmental patterning of archaeological sites have been assumed to be affected the distances to water sources, elevation and the soil conditions of the area, among other factors (Brandt *et al.* 1992, 269). Well-drained, loamy-textures soils are typically best suited for use as agricultural land (Wilcox 2014, 344), and therefore would likely have sites occurring in these areas. Within the Roman period, it is known that soil type was a factor in Roman rural settlement location (Verhagen *et al.* 2014, 382), and therefore may also be the case within Roman Hertfordshire. Therefore, information on the soil textures present within Hertfordshire was an important addition to the model. The original model contained fifteen types of soil textures, but to better represent the soil conditions in the Roman period a new classification scheme was created. A simpler scheme was used which grouped soil textures into five groups, clay, loam, sand, silt and mixed soils. These were grouped by the predominant texture in each original soil class. However, due to the poor resolution of the model version explicit use in the predictive model would be unreliable.

In regards to the preservation of undiscovered archaeological sites, certain soil types are more at risk of large-scale soil excavation or agricultural activity which can dramatically damage or disturb the context of material (Lauwerier *et al.* 2018). The type of agricultural activity can determine the extent of this damage, with annual tillage affecting the top 30cm of the soil (Lauwerier *et al.* 2018). The chemical composition of the soil type can also have an effect on the preservation of certain archaeological materials (Hopkins 2004, 169). For example high levels of preservation can be found in waterlogged, anoxic conditions, whereas low levels of preservation can be present in acidic, sandy soils (Hopkins 2004, 171).

Within Hertfordshire, a majority of the soil has a loamy texture (68%, fig. 9), located in areas of higher elevation. In the modern age, much of the area is located on the chalk escarpment which is known as the Chiltern Hills (Tereszczuk 2004, 9). These calcareous soils were likely formed over millennia by the white chalk bedrock layer from the Upper Cretaceous and influenced by the clay and till superficial layers from the Quaternary period. Much of the soils are deep and

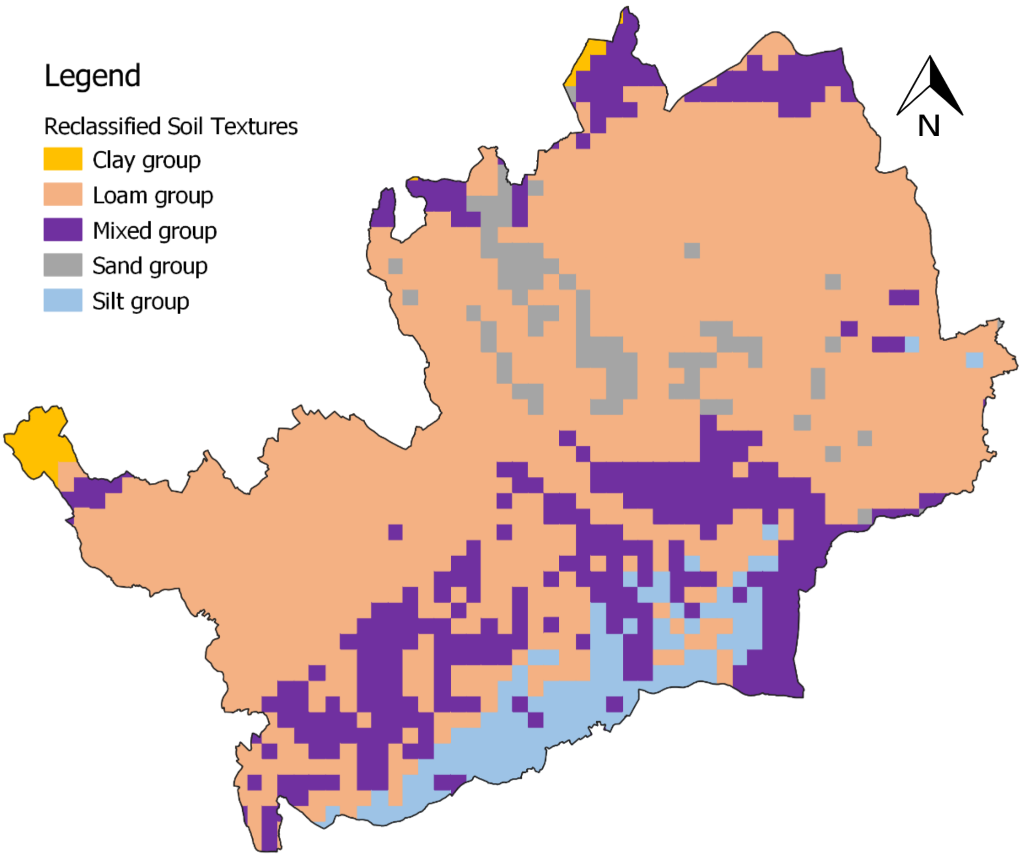


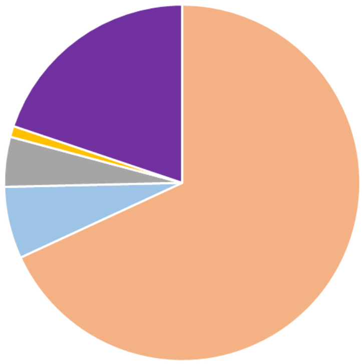
Figure 9: Distribution of reclassified soil textures in Hertfordshire. Based upon the 'Soil Parent Material Model', with the permission of the British Geological Survey.

well drained (Tereszczuk 2004, 8). The mixed soil group is made up of a majority of 'sand to sandy loam' soils and constitutes around 20% (tab. 2) of the total soil textures in Hertfordshire. The small clay group, located mostly at the north-western part of the county, is associated with Jurassic or cretaceous clay and other associated drift (Tereszczuk 2004, 8). The silt group, covering the lower

6.5% of the county (tab. 2), contains clayey material with impended drainage (Tereszczuk 2004, 9).

Table 2: Percentage and area of reclassified soil textures in Hertfordshire.

Soil Texture Groups	Area (km²)	%
Loam Group	1119.614	**68.1%**
Silt Group	107.276	**6.5%**
Sand Group	74.356	**4.5%**
Clay Group	16.395	**1.0%**
Mixed Group	326.117	**19.8%**
Total	1643 km²	100%



3.2.2. Geology

Two bedrock layers were collected for references of soil and elevation contexts, the lower bedrock and the superficial bedrock. Both were provided by the British Geological Survey (BGS) group (www.bgs.ac.uk). These layers are important for understanding the underlying factors of the environment that was present in Hertfordshire in antiquity. The formations which created the bedrock layers influence the elevation, soil composition and waterways that further influence many other factors in the landscape, both environmentally and socially. In some cases, the superficial bedrock layer could be used to predict where industrial extraction sites could have been located archaeologically.

The lower bedrock layer (fig. 10) constitutes the main mass of solid rocks that form the crust of the earth. This is present among the whole of England, fully covering the surface of the island, and is only partially covered by the superficial layers. The ages of the associated formations within Hertfordshire range from the oldest gault formation and upper greensand formation formed in the Early Cretaceous (145-100 Ma) to the Thames group layers which date to the Eocene

(56-33.9 Ma) (appendix 2). The Thames group with marine origins covers the older Lambeth group.

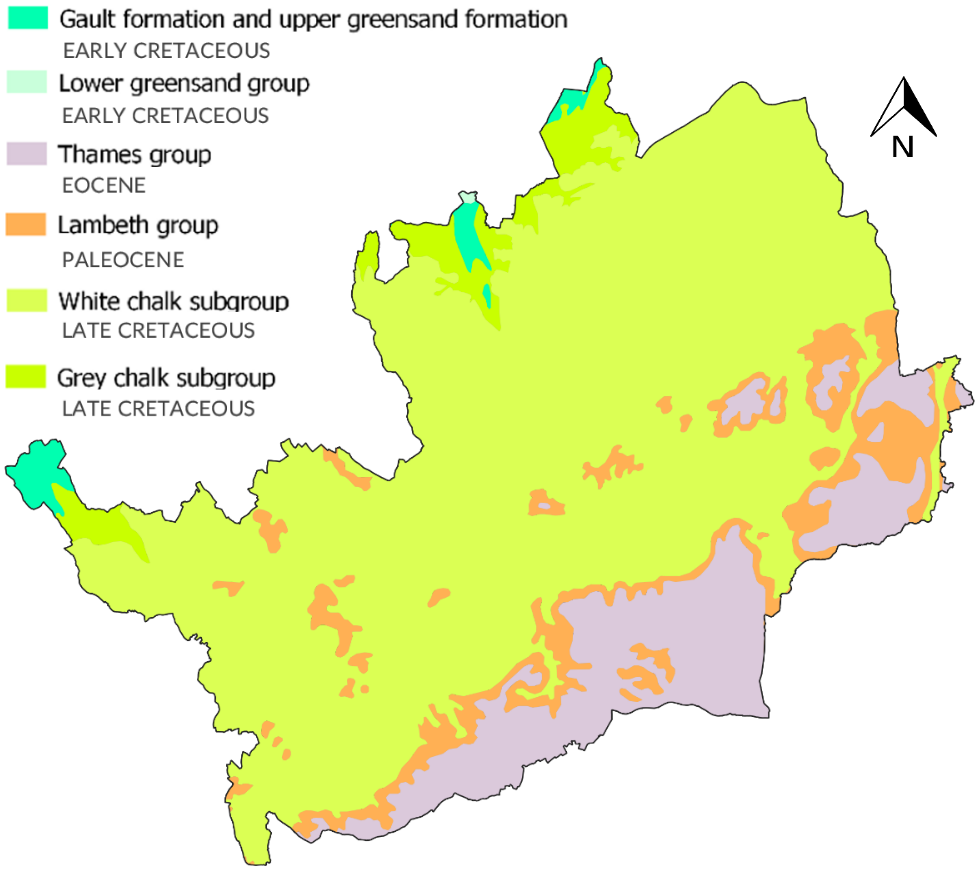


Figure 10: Bedrock geology of Hertfordshire. Based upon the 'BGS Geology 625k' model, with the permission of the British Geological Survey.

The layer of superficial geology (fig. 11) includes the most recent forms of geological deposits, dating to the geological time period, the Quaternary (2.6 Ma) (appendix 3). During this era, the temperature cooled and glaciers covered the middle and north of Britain. Most deposits are shallow, unconsolidated sediments of gravel, sand, silt and clay. Due to the layering of geology, the superficial deposits are the closest to the surface before the soil layer, and only partially cover the lower bedrock in the area of Hertfordshire. Layers of glacial sand and gravel underlie the majority of the mixed soil texture group, likely due to the glaciers depositing an amalgamation of different soil minerals after melting. The alluvial deposits are located where main river channels are located in Hertfordshire, containing clay, silt and sand. The adjacent river deposits contain sand and gravel. The silt soil texture group likely was formed from the lower bedrock deposit from the Thames group as the superficial layers have little coverage in the south of Hertfordshire.

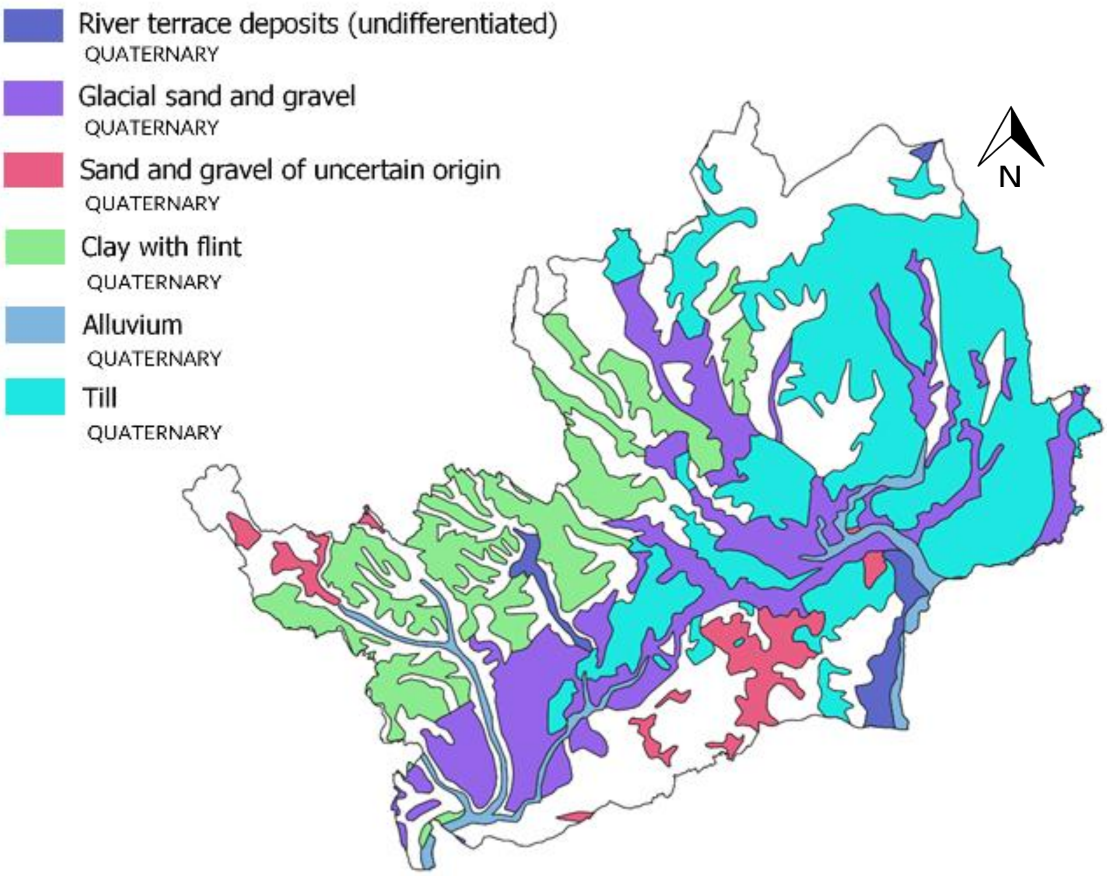


Figure 11: Superficial geology of Hertfordshire. Based upon the 'BGS Geology 625k' model, with the permission of the British Geological Survey.

3.2.3. Hydrogeology

The hydrogeology layer (fig. 12) was also provided by the British Geological Survey (BGS) group (www.bgs.ac.uk). This layer indicated the aquifer potential from geological formations. Other layers were offered by the BGS, such as a water permeability layer, but the hydrogeology layer was instead chosen to be more representative of the groundwater in antiquity (www.bgs.ac.uk).

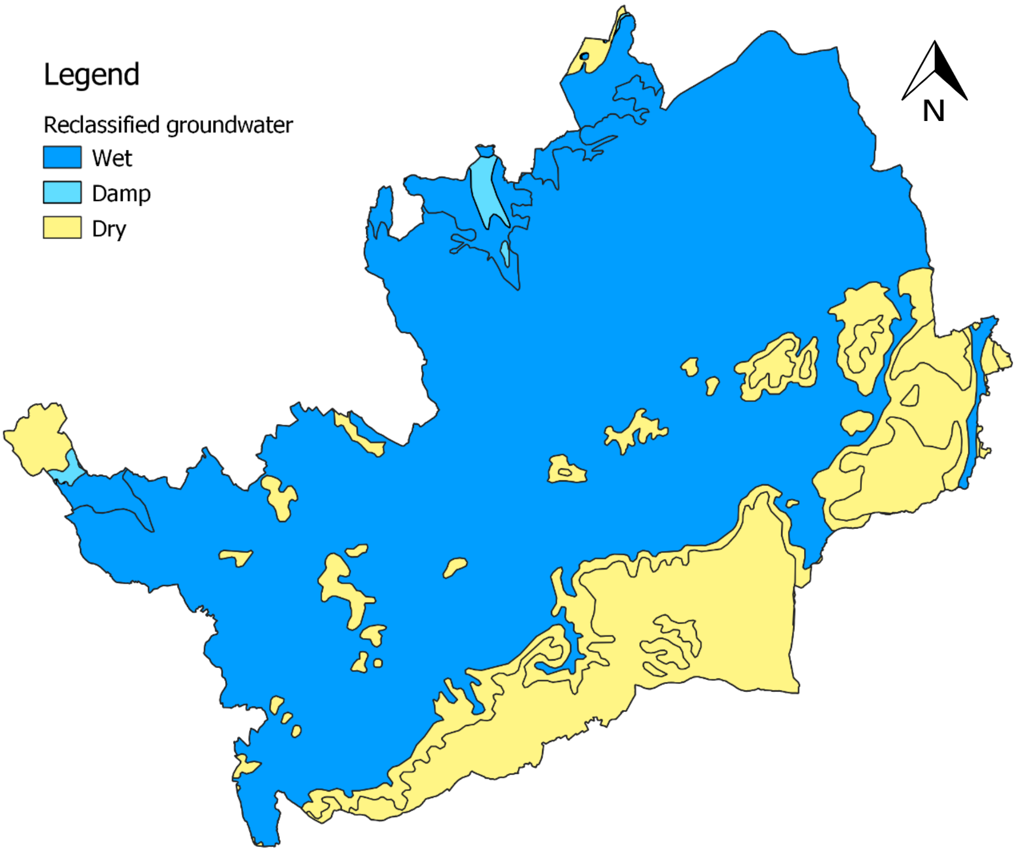


Figure 12: Based upon the 'BGS Hydrogeology 625k', with the permission of the British Geological Survey.

By including this layer, levels of high, medium or low groundwater can be established which may be useful in predicting the locations of archaeological sites as well as the preservation of such archaeology. The "first evidence of extensive drainage in Britain was from the Roman period" (Brown 1997, 269) so drainage systems could have been used to partially control the groundwater levels in parts of the landscape. It is difficult to assume site location based on this data without previously known preferences of low or high groundwater, as both could have been advantageous in site location.

The hydrogeology layer was reclassified to simplify the levels of groundwater to areas that are wet, damp or dry. The underlying superficial and lower bedrock deposits would have influenced the level of groundwater, and the extent of the groundwater would have continuous influence on the soil textures. With this in mind, it can be seen that the areas where loam-textures soils occur is also where the groundwater is almost entirely classified as wet (75%, tab. 3). The groundwater level is dry mostly in the lower southern parts of Hertfordshire where the Thames and the Lambeth group deposits are located (24%, tab. 3). The areas considered 'damp' constitute less than 1% of the area of Hertfordshire, so in terms of groundwater, there are mostly the two extremes of wet and dry.

Table 3: Percentage and area of reclassified groundwater levels in Hertfordshire.

Groundwater level	Area (km²)	%
Wet	1234.09	**75.1%**
Damp	9.61	**0.6%**
Dry	400.05	**24.3%**
Total	1643 km²	100%

3.2.4. River system

A layer showing where water was located in the Roman period was a needed inclusion to the predictive model as water access, or the proximity to water bodies, is one of the main environmental factors that is likely to influence site location in antiquity (Danese *et al.* 2014, 43; Brandt *et al.* 1992, 269). This need was met by the rivers system data, provided by the Ordnance Survey (OS) (www.ordnancesurvey.co.uk). However, rivers are in a constant state of movement and change (Rogers 2013, 89), altering the course by which it takes through the landscape. In order to use this modern river layer for a Roman context, the layer was reclassified into what the main branches were in order to separate them for analysis (fig. 13). Both the elevation layer and the bedrock layers were used to try and determine these older river branches, and much of the original river layer was ultimately used for water proximity analysis.

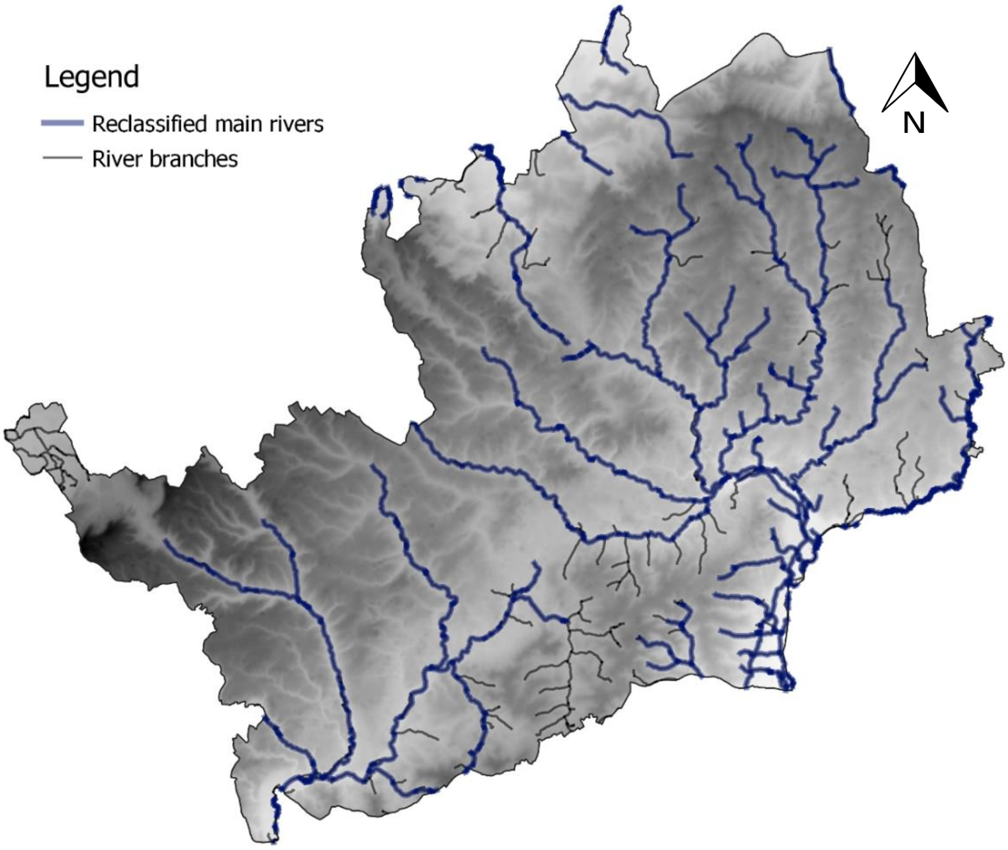


Figure 13: Reclassified main rivers and river branches in Hertfordshire. Based upon the 'OS Open Rivers' layer, with the permission of the Ordnance Survey.

Water is not only needed for human survival, but also directly related to subsistence economies like agriculture. Water has also been known to have cultural values attached to its location and the places through which water has flowed (Rogers 2013, 14). In early antiquity, natural sources of freshwater water, such as lakes, ponds and rivers, were used to fill this need of water access. However, in Roman Britain alternative ways of accessing and controlling water was achieved. Due to this, the supply, distribution and storage of water has formed an important part of Roman urban studies (Rogers 2013, 6). Water mills, man-made channels, canals and wells have been identified in the literature about Roman Britain (Brown 1997, 260; Historic England 2018, 4), as well as in the archaeological dataset in Roman Hertfordshire. Through the deliberate irrigation and drainage of the landscape, areas that humans would have deemed unsuitable for habitation were now able to be settled. This development in

Roman British society would therefore have affected predicted locations for undiscovered sites.

3.3. Roman Roads

Social elements of the landscape can impact site location patterns, with roads included as a main element (Brandt *et al.* 1992, 269; Kamermans *et al.* 2004, 6). Within the Roman era, road networks connected the empire in a scale that was unseen before in antiquity. Within Hertfordshire, it has been deemed that occupation and activity was "clearly influenced by the road" (Fulford 2015, 75). There are also a characteristically large number of roadside settlements within the Roman period which focus on major roads (Historic England 2018, 2). Therefore, the layer of Roman roads that were constructed in Hertfordshire (fig. 14) was used in creating the predictive model of Roman Hertfordshire, with the idea that proximity to these roads would be a factor in site location.

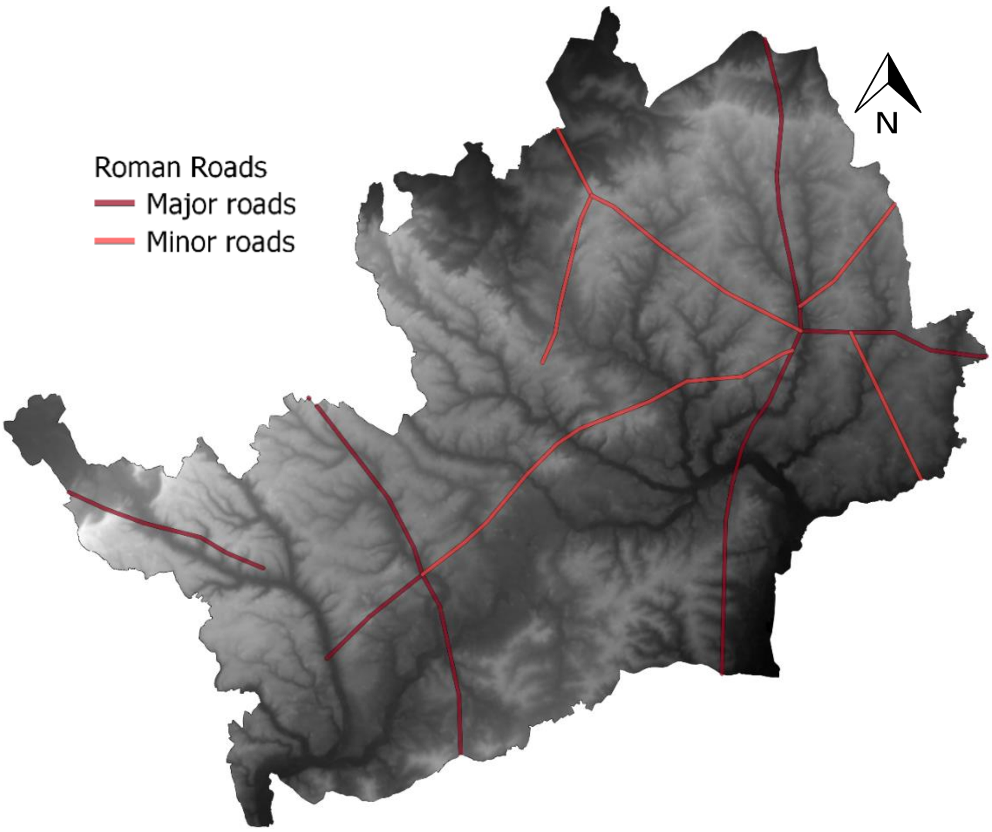


Figure 14: Major and minor Roman roads in Hertfordshire. Based upon the 'Roman Road Network (2008 version)' layer, with the permission of Harvard University.

The data source was created using the 2008 digital version of the Roman roads by McCormick *et al.* (2013), published by Harvard University as a part of 'The Digital Atlas of Roman and Medieval Civilizations' (darmc.harvard.edu). It features both minor and major roads in Roman Britain and across the Roman Empire, based on the '*Barrington Atlas of the Greek and Roman World*' by Richard Talbert, published by Princeton University Press in 2000. A high level of certainty was given to all of the roads that appear within Hertfordshire, according to the data source.

This network of roads linked "developing urban and commercial centers" (Tereszczuk 2004, 10), making the transport links attractive to settlers in the area. People from the North, outside of Hertfordshire, would also pass through this area of Britain while travelling to London (Londinium) along the "main strategic road" of Ermine Street, connecting London with the north of the island (Tereszczuk 2004, 11) (fig. 15). Stane Street was said to have linked centers like Verulamium to Colchester (Fulford 2015, 75), while Watling Street would have linked Verulamium to London.

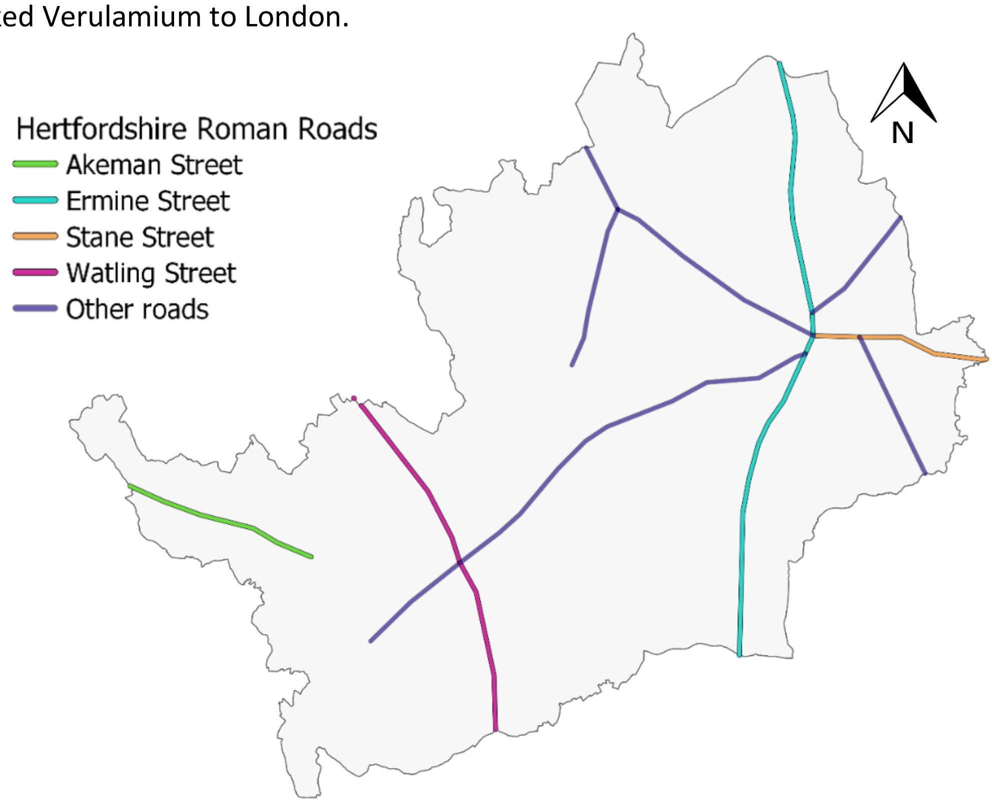


Figure 15: Named Roman streets in Hertfordshire which connected the area to other centers, such as Londinium (London) and Colchester.

3.4. Modern Layers

3.4.1. County and district boundaries

The county of Hertfordshire is made up of ten districts, of which North and East Hertfordshire constitute the largest parts (fig. 16). The district and county boundary polygon data was provided by the Office for National Statistics and hosted on the Open Geography Portal (www.geoportal.statistics.gov.uk). While the county limits of Hertfordshire did not exist within the Roman era, a predictive model suitable for Archaeological Heritage Management (AHM) purposes in mind should be linked to modern contexts. This poses various issues regarding the validity of the assumptions made within the model, as it would then fail to consider factors which occurred outside the modern boundaries of Hertfordshire. However, some limitations must be placed on predictive models by means of its research boundaries.

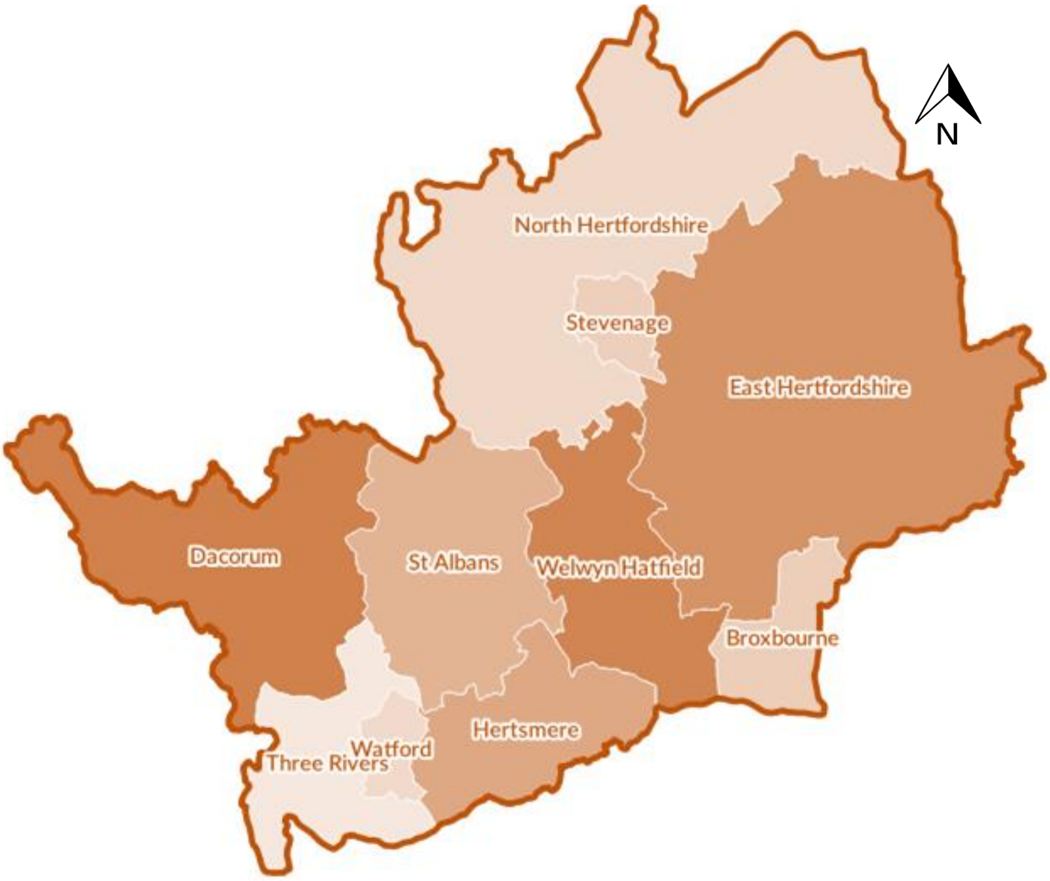


Figure 16: The ten districts within Hertfordshire. Based upon the 'Counties (April 2019) EN BFC' data source, with the permission of ONS Geography Open Data.

The county boundary was used to clip every layer extent to limit the data to the research area of Hertfordshire. However, a layer displaying district boundaries was also needed as the accessible data of legally protected areas were limited to the district of North Hertfordshire. Therefore, reference of where the North Hertfordshire boundary is located was useful for displaying the protected areas.

3.4.2. Modern land-use and roads

Research bias and the state of preservation serve as two factors which can affect the discovery of archaeological sites, both of which are impacted by the modern usage of land. Differences in land use can account for one of the research biases which occurs in search for new archaeology sites, along with differences in survey conditions, collection methods and individual differences (Van Leusen 2002, 76). In regards to the preservation of archaeology, risks can be defined as the product of hazards, vulnerability and exposure (Danese *et al.* 2014, 42). Anthropic hazards is a main risk to the preservation of archaeology, through events like urban sprawl and large-scale infrastructure (Danese *et al.* 2014, 42). Investigations have been conducted on the extent of this risk posed by building foundations, finding that the load-bearing layer is often the same which contains archaeological remains (Bouwmeester *et al.* 2017, 150). It is therefore crucial that areas of different modern land use are known in order to account for both the biases and archaeological risks.

It is for this reason that the modern land use layer, provided by the European Copernicus Land Monitoring Service (ECLMS) (www.land.copernicus.eu), was reclassified to best serve as a basis for the five main types of land use occurring in Hertfordshire (fig. 17). It can be seen in this layer that much of the land is used as cropland (72.4%) and has a significant level of urban sprawl (21.4%) (tab. 4).

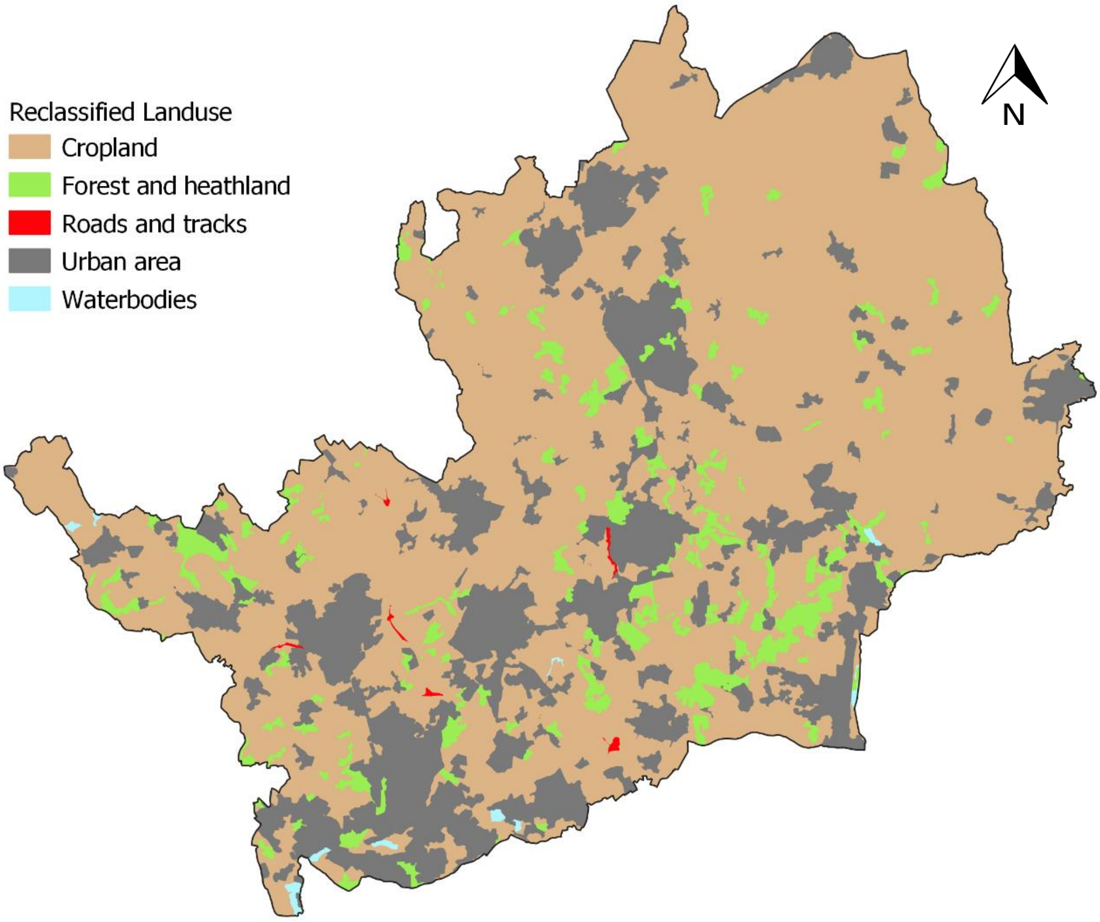
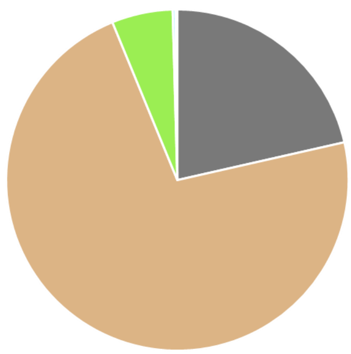


Figure 17: Reclassified modern land-use in Hertfordshire. Based upon the 'Corine Land Cover (CLC) 2018' data source, with the permission of the Copernicus Land Monitoring Service.

Table 4: Percentage and area of modern land-use in Hertfordshire.

Modern Landuse	Area (km²)	%
Urban area	352.42	**21.4%**
Cropland	1189.12	**72.4%**
Forest and heathland	94.51	**5.8%**
Waterbodies	4.52	**0.3%**
Roads and tracks	2.79	**0.2%**
Total	1643 km²	100%



A layer displaying the modern roads constructed in Hertfordshire was provided by the Ordnance Survey (OS) (www.ordnancesurvey.co.uk). This was used to both add navigational references to the predictive model, as well as to provide additional information on the preservation of unknown and known archaeological sites (fig. 18). Commonly, predictive models have been used for "large-scale highway planning purposes" (Verhagen & Whitley 2012, 54; Podobnikar *et al.* 2001, 544) in other countries, and therefore references of existing roads may indicate to developers the suitable areas for highway development which do not require expensive archaeological research.

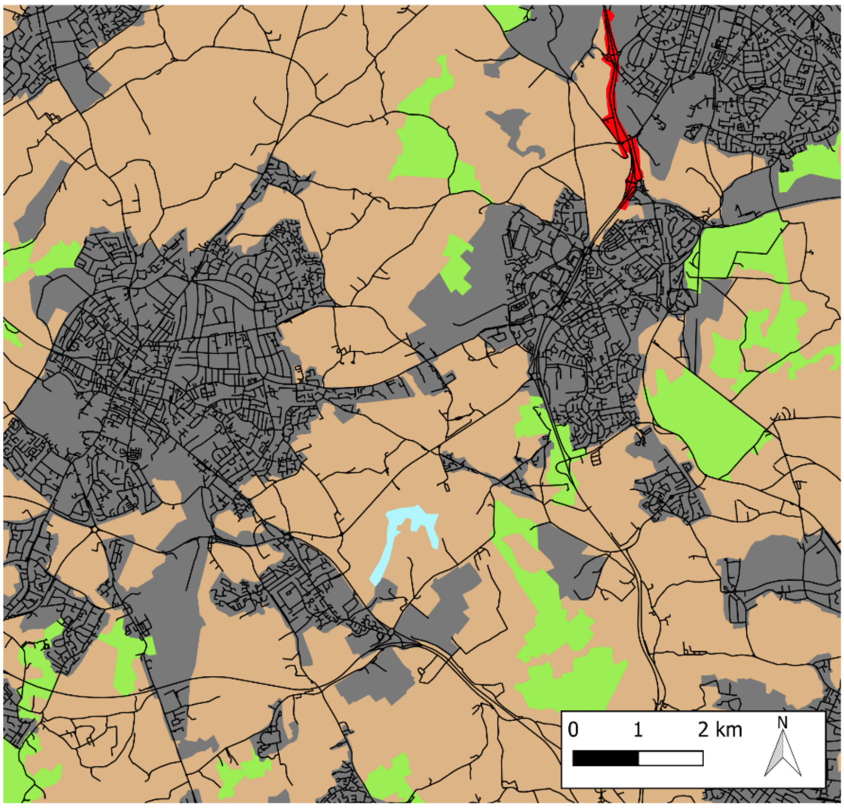


Figure 18: Placement of modern roads around modern land-use in a part of Hertfordshire. Based upon the 'Corine Land Cover (CLC) 2018' data source, with the permission of the Copernicus Land Monitoring Service, and the 'OS Open Roads' layer, with the permission of the Ordnance Survey.

3.4.3. Protected areas and scheduled monuments

Local protections on archaeology and historical monuments are important elements to include in the display of an archaeological predictive map. Within England, planning permissions for development projects are often granted through the decisions of the Local Authority (www.archaeologists.net; English Heritage 2015, 1) and therefore a rudimentary knowledge of the areas where any kind of development is not possible would likely save time and money. Two layers were collected to display these protected areas: a layer of scheduled monuments which was provided by Historic England (historicengland.org.uk) and data on the 'Archaeological Areas' in North Hertfordshire, provided by the North Hertfordshire District Council (www.data.gov.uk) (fig. 19).

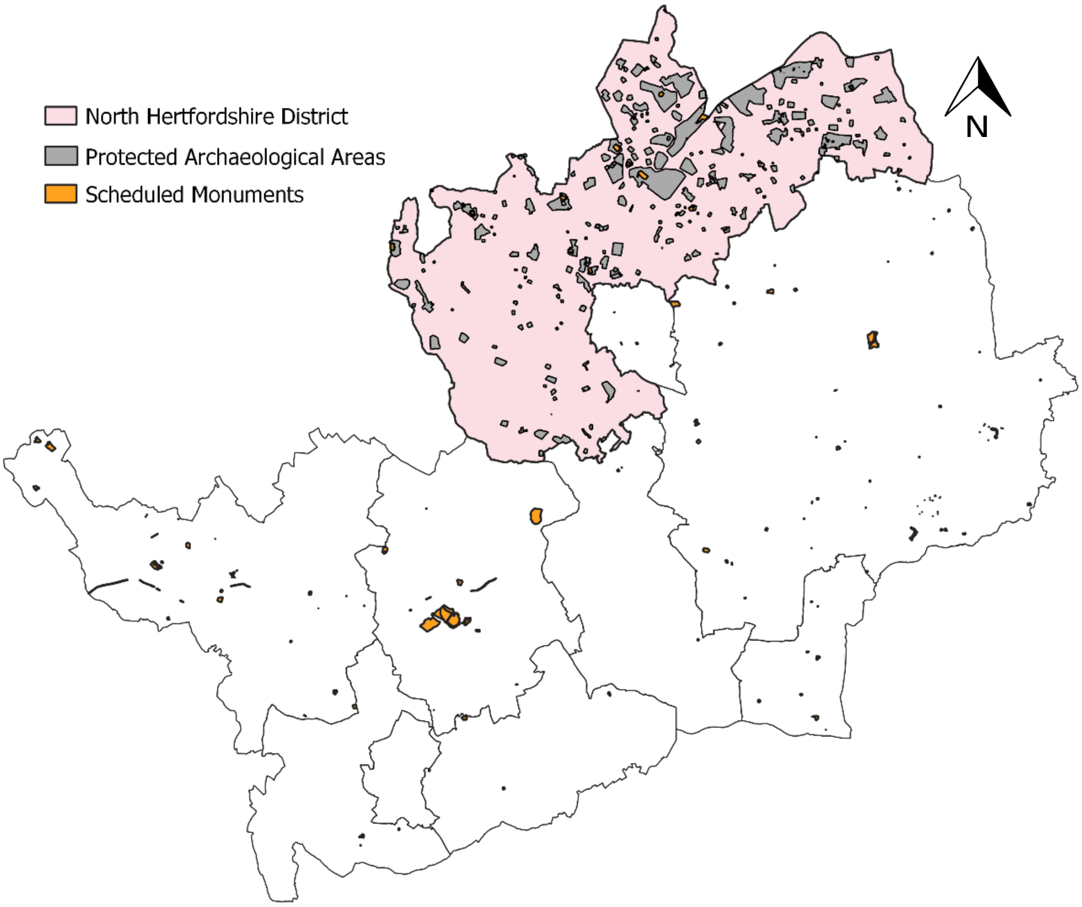


Figure 19: Archaeological areas (in North Hertfordshire) and scheduled monuments in Hertfordshire. Based upon the 'Archaeological Areas' data source, with the permission of the North Hertfordshire District Council, and the 'Scheduled monuments' layer, with the permission of Historic England.

The scheduled monuments layer displays the size and location of any kind of "nationally important archaeological sites" within the boundary of Hertfordshire (historicengland.org.uk). Scheduling is the oldest form of heritage protection in England, originating from the 1882 Ancient Monuments Protection Act. The layer includes monuments that date before and after the Roman period so that there is awareness for all protected monuments in the area.

The layer of 'Archaeological Areas' are designated as such for their archaeological importance under the Ancient Monuments and Archaeological Areas Act 1979 (www.data.gov.uk). Unfortunately, the archaeological areas from the other districts of Hertfordshire could not be found, so the protected archaeological areas were only used as accompanying background information, rather than directly informing the predictive model to ensure consistency.

3.5. Archaeological Data

The archaeological data which was used for the predictive model of Roman Hertfordshire came from the open-access archaeological database, the Archaeological Data Service (ADS). Since the organisation's establishment in York, England in 1996, the ADS has become "the United Kingdom's national digital data archive for archaeology", and is the "longest serving repository for archaeological data in the world" (Richards 2017, 227). It holds archaeological records that can be used spatially through its inclusion of coordinates, as well as archaeological literature which are collected by various heritage agencies in the UK (Wright & Richards, 2018, 61-62). The 'ArchSearch' function on the main ADS website is an "integrated online catalogue" (www.archaeologydataservice.ac.uk) that allows the searching of over 1.3 million metadata records from over 30 historical inventories (Wright & Richards, 2018, 62). Its function to query data on the basis of site type, site location, site date and archaeological source enables the selection of specific data for re-analysis.

The ADS database stores a total of 1352 records within Hertfordshire which date to the Roman period. A request was made to employees of the ADS team in order to receive the full result of the query of Roman sites within Hertfordshire. This was necessary to by-pass the sample-limit they impose on site users.

3.5.1. Data cleansing

After receiving the requested data in the form of a CSV (Comma-Separated Values) file, the data was imported into Microsoft Excel. A measure of data cleansing was necessary before it could be imported into QGIS (fig. 20). This included separating data into more individual columns. In order to use the data spatially within QGIS, X and Y coordinates needed to be separated into individual columns. However, additional columns were also made for the attributes 'Named Location', 'Grid References', 'Civil Parish', 'District', 'Subject', 'Bibliographic References', 'URL', 'Depositor ID', 'Creator' and 'Publisher'. The fields containing the 'Title' and 'Description' were already separated in the original format.

	C	
1	Location	Period
2	Grid Ref:TQ1647098900;Civil Parish:ALDENHAM;EPSG:27700:198900;District:HERTSMERE;Country:England;Ad	Subject
3	Grid Ref:TQ1647098900;Civil Parish:ALDENHAM;EPSG:27700:198900;District:HERTSMERE;Country:England;Ad	Interve
4	Named Location:'Fourways' Cannon's Close;District:EAST HERTFORDSHIRE;EPSG:27700:222094;Admin County	Subject
5	EPSG:27700:205875;EPSG:27700:503750;Country:England;Admin County:HERTFORDSHIRE;District:DACORUM;	Interve
6	EPSG:27700:524530;Civil Parish:BALDOCK;Grid Ref:TL2453034080;Country:England;Admin County:HERTFORD:	Interve
7	EPSG:27700:223580;District:EAST HERTFORDSHIRE;Civil Parish:STANDON;Country:England;Admin County:HEI	Interve
8	Civil Parish:LETCHWORTH;EPSG:27700:521540;Country:England;Admin County:HERTFORDSHIRE;EPSG:27700:2	Interve
9	Named Location:124-126 High Street;District:STEVENAGE;Admin County:HERTFORDSHIRE;Country:ENGLAND	Period:
10	District:EAST HERTFORDSHIRE;Grid Ref:TL4946021940;Civil Parish:BISHOP'S STORTFORD;Country:England;Ad	Subject
11	District:EAST HERTFORDSHIRE;Grid Ref:TL4948021960;Civil Parish:BISHOP'S STORTFORD;Country:England;EPS	Interve
12	EPSG:27700:525420;Named Location:Westell Close	Baldoc
13	Civil Parish:WYMONDLEY;EPSG:27700:228600;Country:England;Admin County:HERTFORDSHIRE;District:NORT	Interve
14	Grid Ref:TL2540033850;EPSG:27700:525400;Country:England;EPSG:27700:233850;Admin County:HERTFORDSH	Subject
15	EPSG:27700:517800;EPSG:27700:195500;Named Location:High Street	Elstree
16	Grid Ref:TL2440024770;EPSG:27700:524400;Civil Parish:STEVENAGE;EPSG:27700:224770;Country:England;Adm	Period:
17	Civil Parish:BALDOCK;Country:England;Admin County:HERTFORDSHIRE;District:NORTH HERTFORDSHIRE;EPS(	Interve
18	District:EAST HERTFORDSHIRE;EPSG:27700:222100;Civil Parish:BISHOP'S STORTFORD;Country:England;Admin	Interve
19	EPSG:27700:524670;EPSG:27700:234050;Civil Parish:Baldock;District:North Hertfordshire;Country:England;Ac	Period:
20	EPSG:27700:524000;Civil Parish:BALDOCK;EPSG:27700:233000;Country:England;Admin County:HERTFORDSHII	Subject
21	Grid Ref:TL2445034050;EPSG:27700:524450;EPSG:27700:234050;Civil Parish:BALDOCK;Named Location:Orchar	Baldoc
22	EPSG:27700:524450;EPSG:27700:234050;Civil Parish:BALDOCK;Named Location:2 - 24 Orchard Road ;Grid Ref:T	Subject
23	Civil Parish:Baldock;EPSG:27700:524750;Named Location:20-22 ICKNIELD WAY EAST;EPSG:27700:234250;Distri	Period:

Roman_Hertfordshire(3062)

Figure 20: Sample of the 'Location' column in the original CSV file export from the Archaeology Data Service, displaying the mix-up of attributes.

The data was imported into Microsoft Access where an auto-numbering ‘ID’ was generated for all 1352 records. Queries were written for each newly created field using the ‘InStr’ function. This generated the location of the data that was to be placed into a new field. To determine the location of the data which was in no specific order, the query aimed to return the position character which marks the beginning of the data and the character length of the data (fig. 21).

Named Location ‘Start’:

InStr(dataromancsv.Location,"Named Location:")+15

Named Location ‘Length’:

InStr(Mid(dataromancsv.Location,InStr(dataromancsv.Location, "Named Location:")+15,Len(dataromancsv.Location)),";")-1

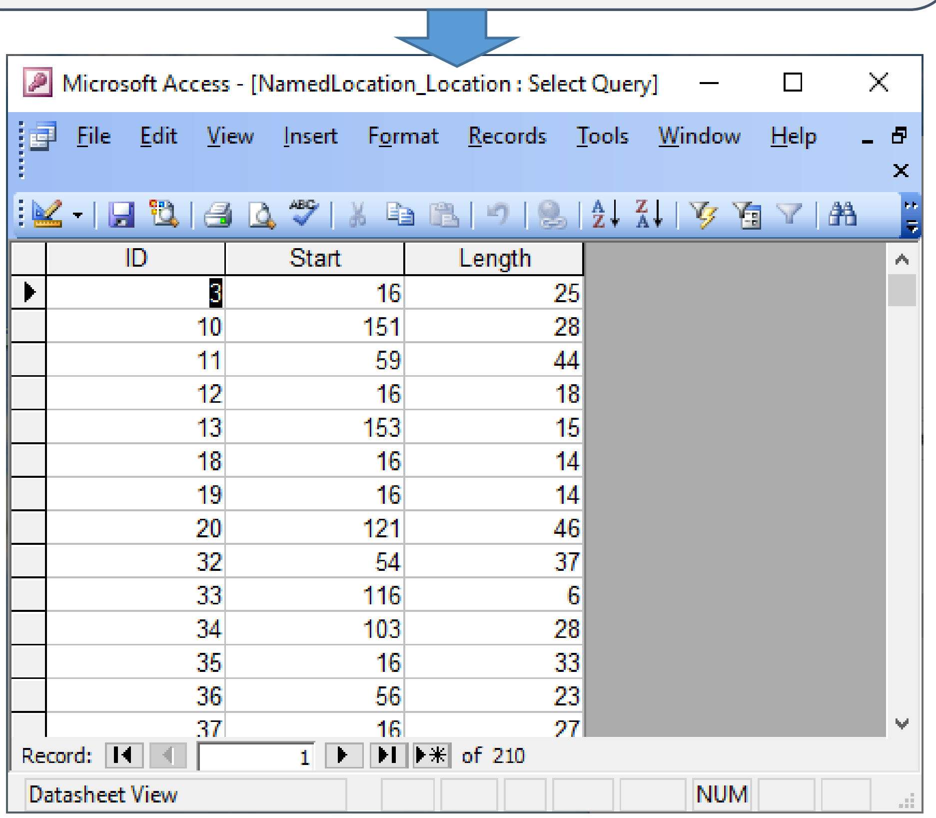


Figure 21: Example of the queries written in Microsoft Access to obtain the location of specific data through the starting character position and character length.

All coordinates listed in the data used the ‘British National Grid’ coordinate reference system (EPSG: 27700), but assigning the correct X and Y coordinate to their respective fields added a level of difficulty in the data cleansing process as the difference was not initially clear. To address this, the two coordinates were first entered into place-holder fields, named 'NUM1' and 'NUM2', in their original order in the data.

After observing the coordinates which were correctly located in Hertfordshire (fig. 22), both the 'NUM1' and 'NUM2' fields were sorted by those greater than 400,000 (as X coordinates) and those lesser than 400,000 (as Y coordinates).

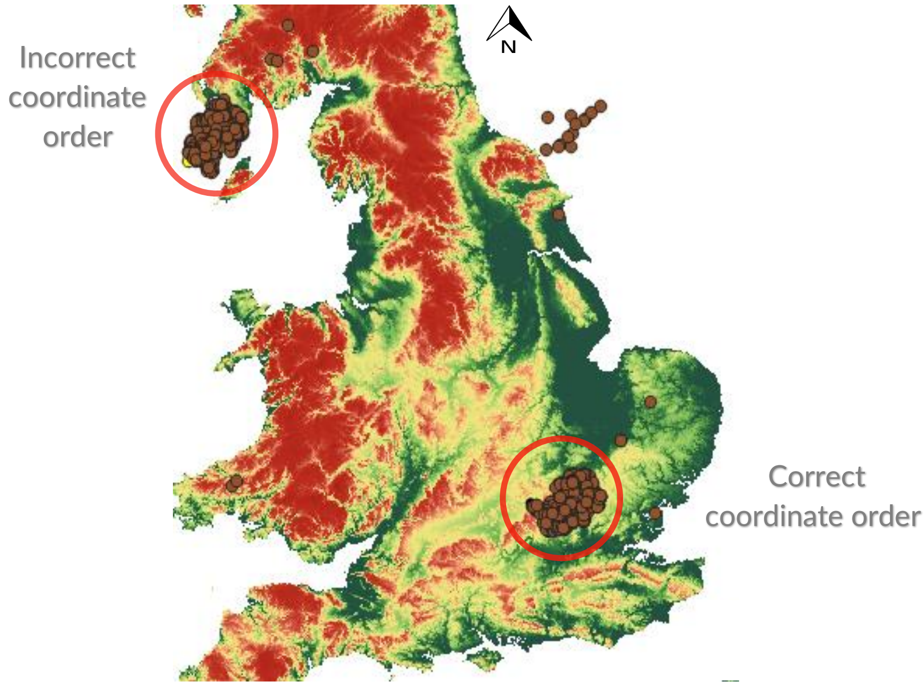


Figure 22: First spatial importation of site points within QGIS, using the original order of coordinates found within the original ADS data.

Once the data was sufficiently organised, the full dataset was imported into QGIS and remaining anomalies and outliers were looked for. This cleaning process removed 41 records, all of which either were outside the Hertfordshire boundaries or did not have coordinate data. Therefore, the remaining 1311 Roman archaeological records now were organised and spatially checked.

Within the remaining data, the 'Subject' field still contained multiple entries for each record. This was likely to be due to the way the ADS stores multiple finds within a single site or excavation. For the purposes of categorising the site types and analysing them separately from each other, records with multiple subject entries were duplicated in order to achieve only one 'subject' per record. The duplicated records were not assigned new unique ID numbers, keeping the original in its place. This was done so that duplicate records could be grouped together, if needed. This increased the number of records from 1311 to a total of 4358 records, however this did not add points with new coordinates.

3.5.2. Split sampling data

The archaeological data was partitioned randomly into two groups to ensure an unused 'test group' was available to estimate the accuracy of the final product. Therefore, the model was 'trained' with around 80% of the data while the remaining 20% of data was used to cross validate the assigned areas of high and low predictive values. In order to decide upon the ratio of training and testing data, two competing concerns were considered. With less training data it is possible that less variance would be visible during the creation of the predictive model and thereby weaken its accuracy and detail. However, with less testing data the capabilities of the sample to determine the performance of the predictive model may be rendered unrepresentative. Due to these concerns, achieving a proportionate balance between the two variances was important.

The final ratio of 80:20 was decided upon due to the rule stated by the Pareto principle. The Pareto principle, also known as the 80/20 rule, states that "a small number of causes (20%) is responsible for a large percentage (80%) of the effect" (Lipovetsky 2009, 271) and therefore such a ratio would produce a representative sample of data. The principle was initially coined in relation to distributions found within economics but heuristically became related to other distributions in life, such as wealth, crime and eventually mathematics (Lipovetsky 2009, 276). The 80/20 division to create a training and a testing sample is also commonly used in the field of machine learning for archaeological site detection (Verschoof-van der Vaart *et al.* 2020, 293). In addition to this, the total number of data records was large enough to confidently remove 20% of data from the creation of the model and remain sufficient. The Pareto principle was therefore deemed appropriate to proportionately divide my archaeological data into a training and testing group.

The ID number, which was generated during data cleansing, was used within a selection query using Microsoft Access. By using the assumption that all of the ID values will end in any number between 0 and 9, we can separate 80% sample of the data if we only select those ID records which end in a number larger than 1 (2, 3, 4, 5, 6, 7, 8, 9). Therefore, the remaining 20% of the data will have ID numbers which end in any number equal or less than 1 (0, 1). This selection process was expressed within the queries seen in Figure 23.

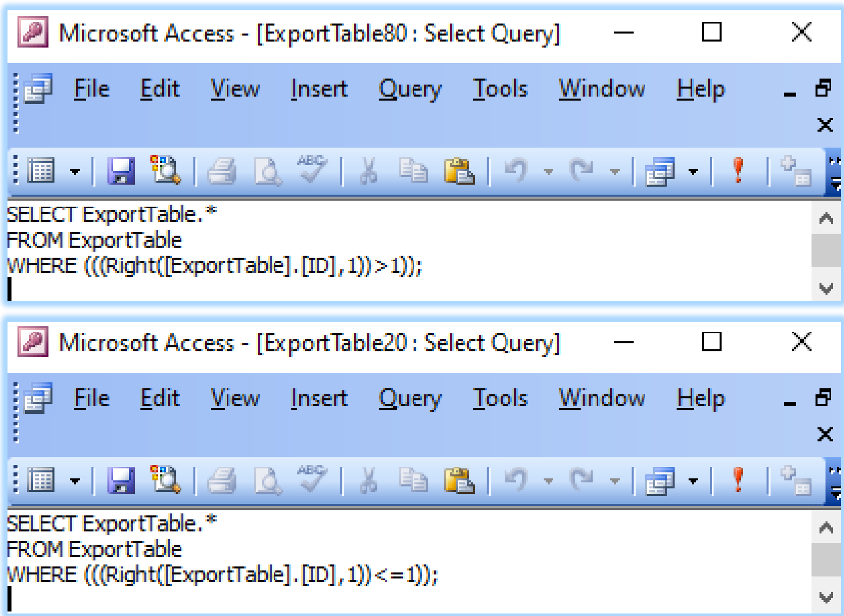


Figure 23: The two selection queries which were used to indiscriminately select 80% of the archaeological data by selecting records with an ID that ends in a number above 1.

This method succeeded in separating the data as randomly as possible by indiscriminately selecting auto-incrementing ID values which were given to unsorted data with no specific ordering. However, instances of duplicate ID values meant that the training and test samples could not amount to exactly 80% and 20%. This was deliberately done so that records which were duplicated earlier in the process remained within the same group. When ‘new’ duplicate records were generated for the multiple “Subject” entries within each original record the duplicate records kept their original ID values. Therefore, some ID values occurred more frequently than others, but still essentially belonged to one single record. Despite this, the method of selecting by ID ultimately assigned the training group 79.5% of the data (*n* = 3466), and the testing group 20.5% (*n* = 892) (fig. 24).

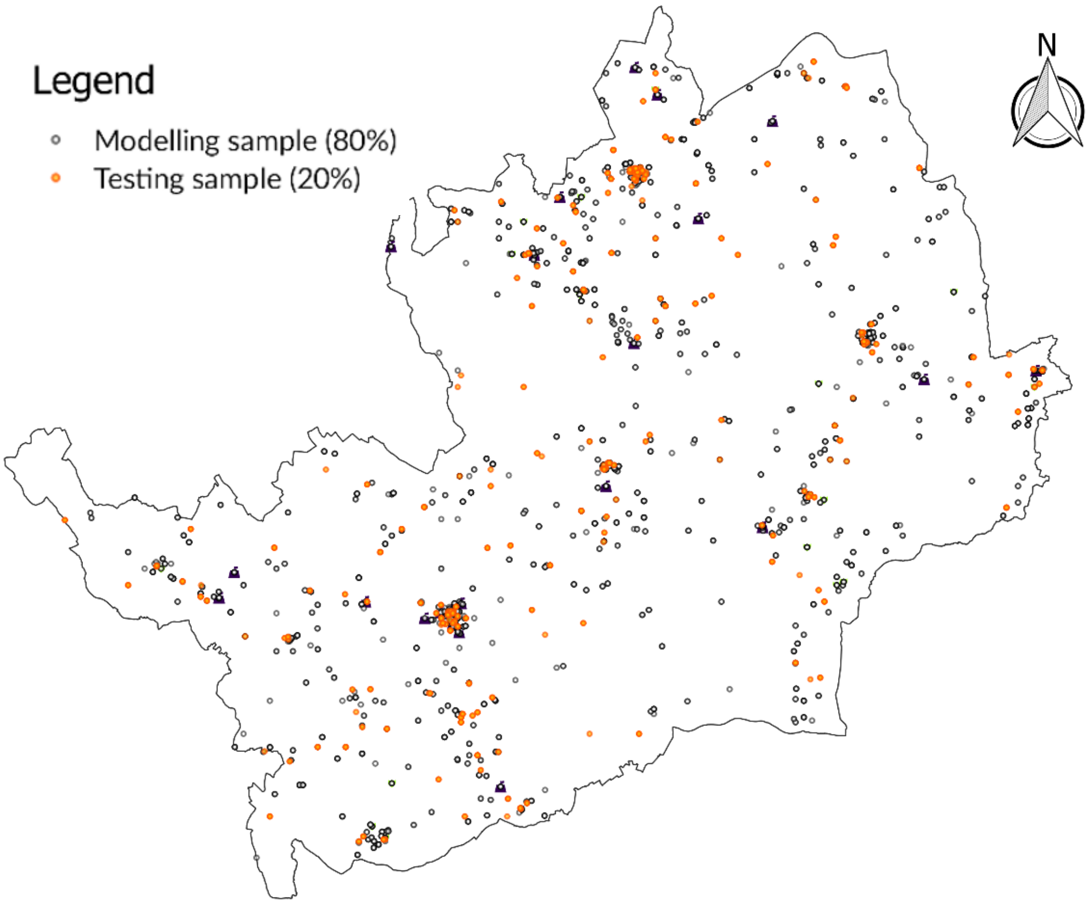


Figure 24: Distribution of the model data sample (n = 3466) and testing sample (n = 892) displayed within the boundaries of Hertfordshire.

3.5.3. Categorising subjects

Categorising the 80% sample of data was crucial if any observations were to be made on the basis of site type. For example, if the proximity to water was to be investigated for influences of site location, the nature of the site would likely affect how much water was needed there. At a Roman ritual site, the main activity which would have taken place there would have been worshipping, with few people residing there permanently. For this reason, landscape characteristics other than the proximity to water could have played a more important role in deciding the site location. Alternatively, settlement sites would theoretically place a high importance of proximity to water sources and may therefore need to be investigated separately from other site types. To do this, these site categories would have to be pre-decided on the basis of the 360 unique 'subjects' within the data.

Categorising the different subjects began with creating a list of unique values within the 3466 records. This was done using the vector analysis tool 'List unique values' in QGIS. This list was imported into Microsoft Excel where each subject was individually placed into one of six groups: settlement, economic, military, ritual, water sources or miscellaneous.

The settlement group (fig. 25) contained subjects which had identified settlement structures such as baths, towns and villas (appendix 4), settlement-related objects such as amulets, ceramics and beads (appendix 5), infrastructure such as enclosures, bridges and roads (appendix 6), as well as unknown structure remains such as bricks, mosaics and architectural fragments (appendix 7).

The economic group (fig. 26) contained subjects related to agricultural processing like corn driers, mills and ovens (appendix 8), agricultural land such as barns, farms and vineyards (appendix 9), industrial production such as quarries, workshops and kilns (appendix 10), in addition to industrial objects like lithic material, coin moulds and crucibles (appendix 11).

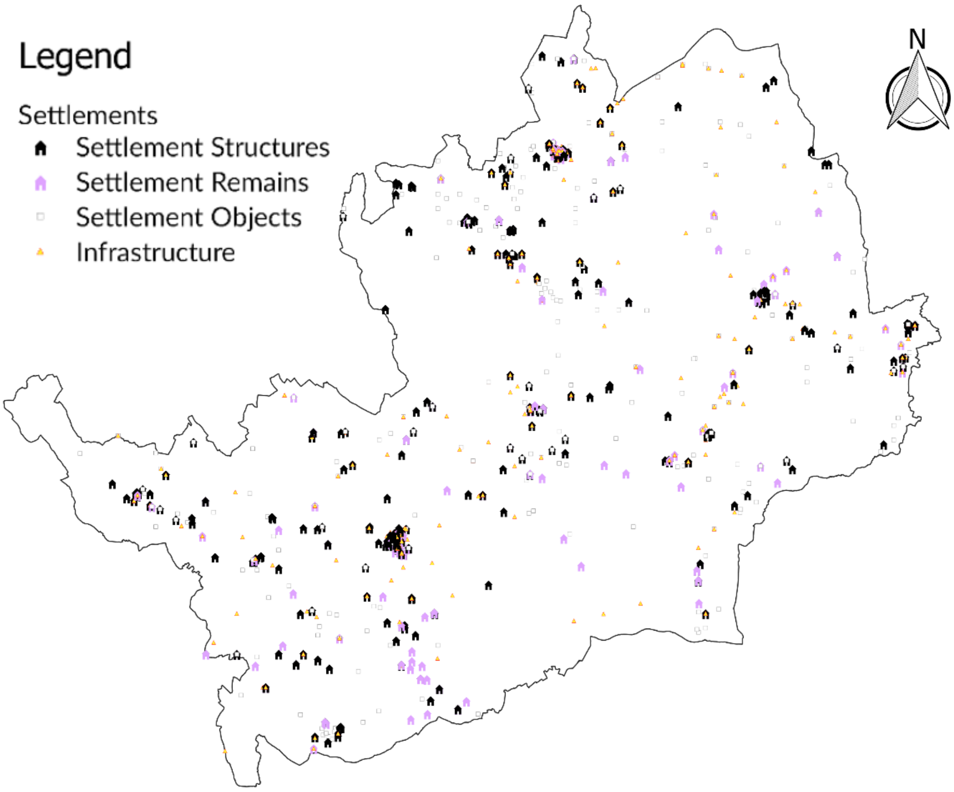


Figure 25: Settlement sites (n = 2017), as classified in appendices 4-7.

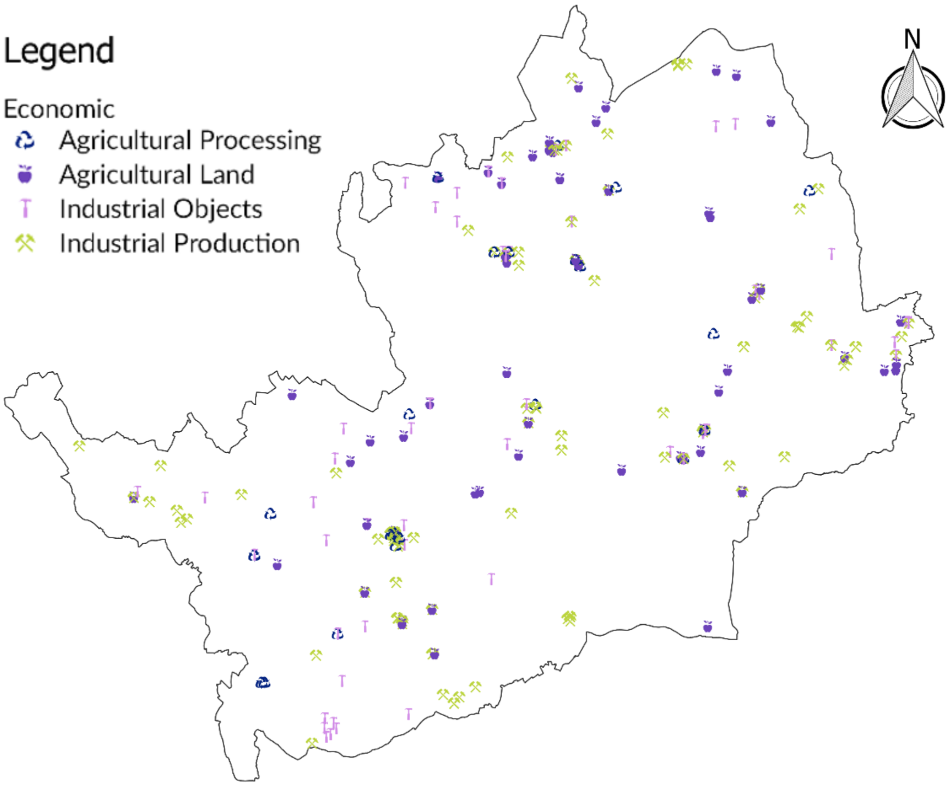


Figure 26: Economic sites (n = 320), as classified in appendices 8-11.

Within the military group (fig. 27) were subjects related to military structures such as forts, moats and towers (appendix 12) as well as military objects like axes, horseshoes, arrowheads and weapons (appendix 13).

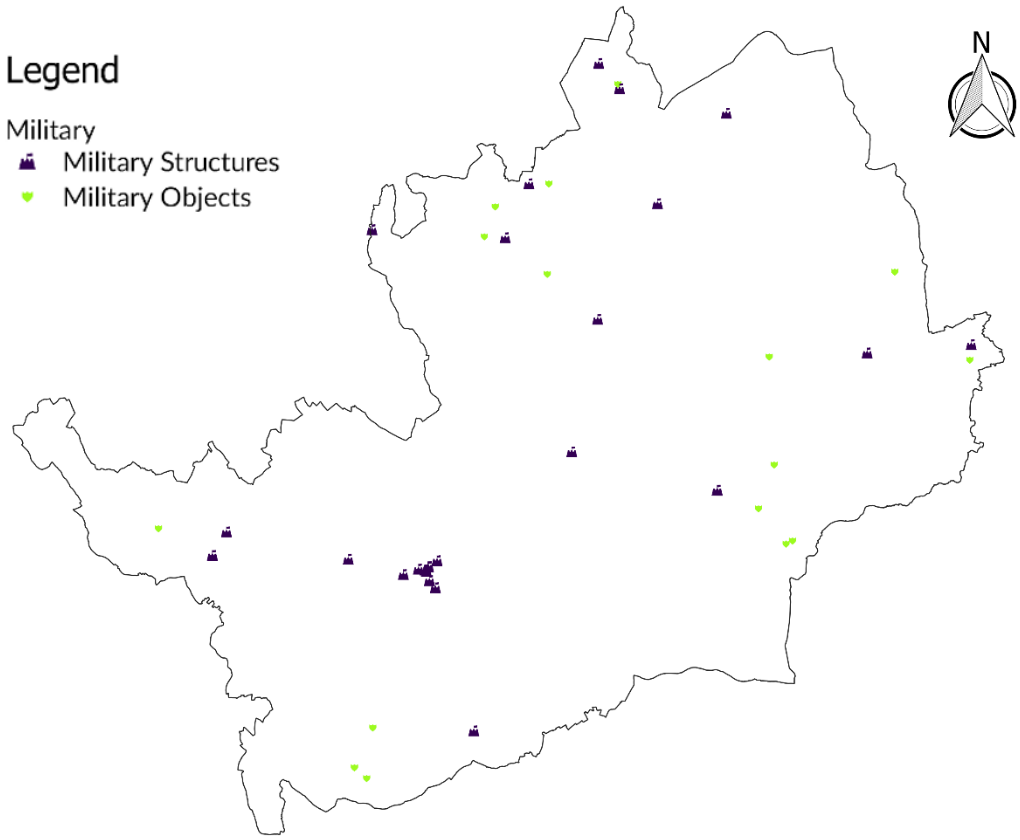


Figure 27: Military sites (n = 53), as classified in appendices 12 and 13.

The ritual group (fig. 28) contained subjects that were grouped as religious structures included ritual pits, shrines and temples (appendix 14), as well as funerary sites such as cremations, urns and mausoleums (appendix 15).

The group of water sources (fig. 29) included subjects which were related to Roman water management, such as arched brick culverts, canals and drainage ditches (appendix 16).

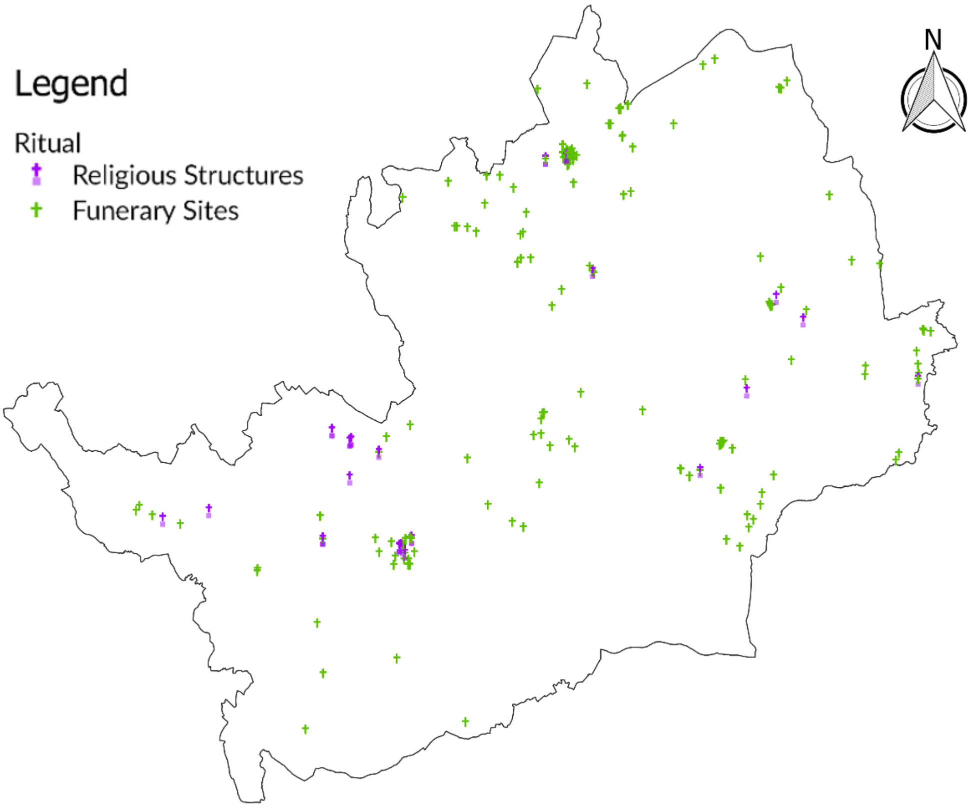


Figure 28: Ritual sites (n = 318), as classified in appendices 14 and 15.

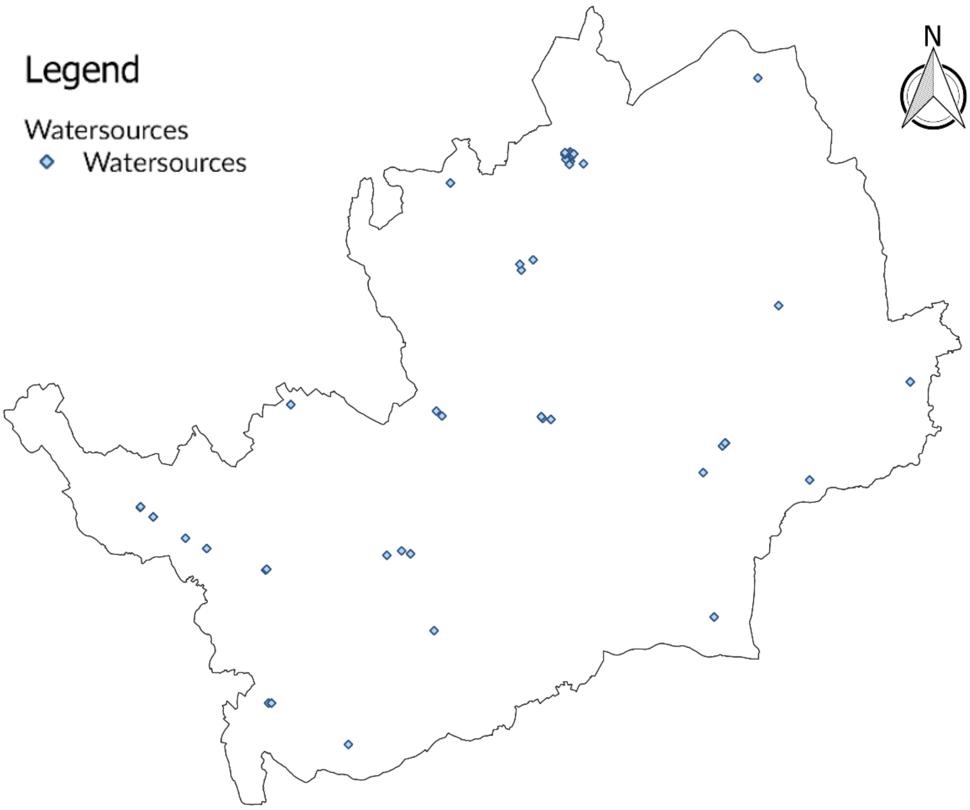


Figure 29: Water source sites (n = 62), as classified in appendix 16.

As for the group of miscellaneous subjects, this category contained both unreliable and uncategorised sites (fig. 30). It appeared that most of the subjects within this group were duplicate points within the same records as other subjects that were more descriptive than being sites, layers or finds – these were categorised as uncategorised subjects (appendix 17). Meanwhile, the unreliable observations included subjects that likely dated to later than the Roman period, such as air raid shelters, Methodist chapels and motte-and-bailey castles (appendix 18).

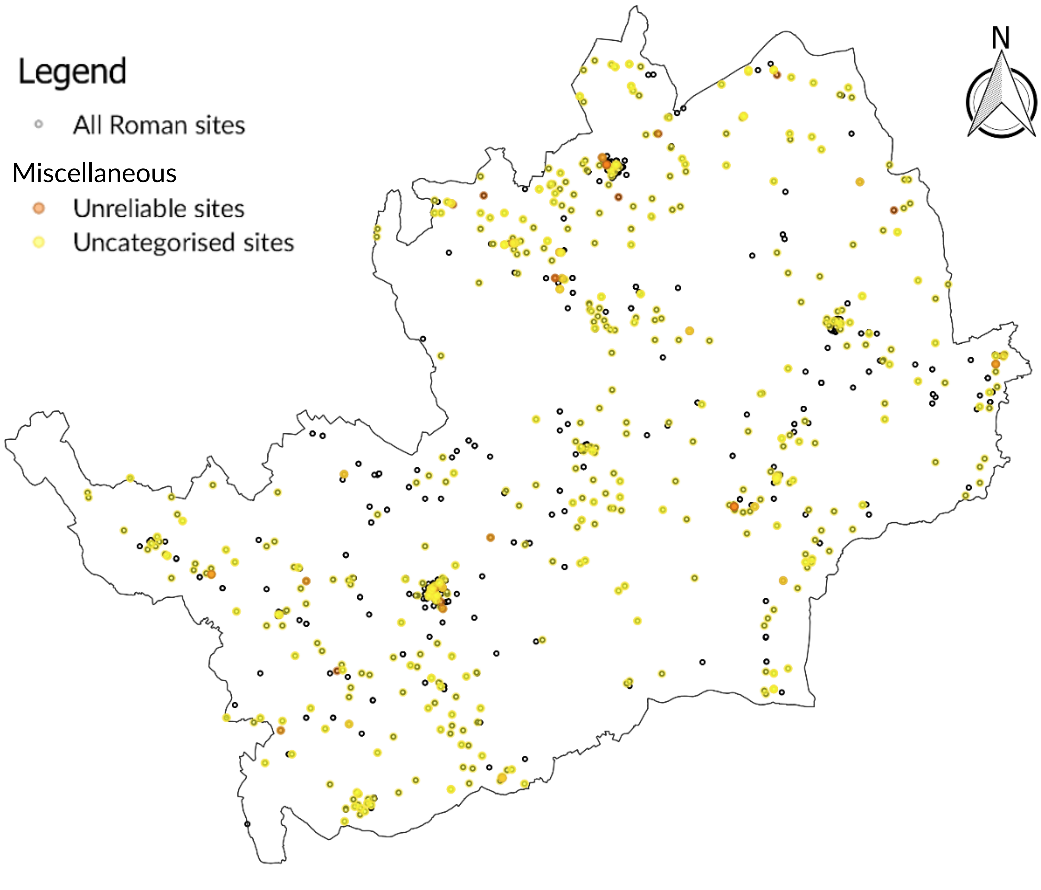


Figure 30: Miscellaneous sites (n = 696), as classified in appendices 17 and 18.

4. Methodology

A detailed explanation of the methods and theories used in the creation of a predictive model is crucial for the reproducibility, critique and use of the model (Wilcox 2014, 340). Therefore, the different methods that were used to create the Roman Hertfordshire predictive model will be explained in this chapter, and the resulting application of these methods will be presented in the 'Results' chapter.

Deductive modelling offers the modeller the opportunity to predict sites on the basis of theoretical locational rules (Danese *et al.* 2014, 43) which are based on archaeological knowledge or theory. It is seen by some that using a deductive, theory-driven framework is a better way to model than the alternative of inductive models (Verhagen *et al.* 2007, 203). Inductive, or correlative, models are created through the extrapolation of locational rules which are derived from patterns within a dataset of known sites. The resulting model from this technique can therefore lack an external testing mechanism (Kamermans *et al.* 2004, 5) and often exclude the influence of social factors due to their difficulty to model (Verhagen *et al.* 2007, 204). Therefore, the Roman Hertfordshire predictive model was primarily created using a deductive approach, which forms predictions on the basis of "prior anthropological and archaeological knowledge" (Kamermans *et al.* 2004, 5). Inductive modelling, based on site density was also used, but only minimally impacted the final model.

4.1. Predictive Factors

The predictive model implemented both environmental and social factors which could have affected the choice of site location and can indicate areas where settlement was highly suitable during the Roman period in Hertfordshire.

The environmental factors that were considered included the proximity from water sources, which accounted for the main river courses within Hertfordshire as well as any previously-identified Roman sources of water access such as aqueducts, pipes or wells. Areas in the closest proximity to both Roman water sources and main rivers (within 1km) were given a 'Very High' archaeological prediction and areas over 5km from water were assigned a 'Very Low' predictive value. Other environmental factors were used to assign a 'Very High' archaeological prediction, which had both a southern-facing aspect and a shallow slope below 10 degrees.

The social factors that were integrated into the model's site predictions included the proximity from the Roman-built road networks that travelled through the boundaries of Hertfordshire and connected the geographical region to other parts of Roman Britain. The wider areas in the closest proximity to the roads (within 1km) were given a 'Very High' archaeological prediction, whereas areas over 5km away from the roads were given a 'Very Low' prediction. Lastly, site densities of known archaeological data were analysed in order to identify where potential major and minor Roman towns were located. This analysis was integrated as 'Very High' areas of archaeological prediction.

4.2. Modelling Methods

4.2.1. Deductive methods

The deductive mode of modelling was implemented through the use of multi-criteria analysis. This type of analysis allows for the inclusion of multiple factors which are presumed to have had impacted Roman site location through the use of a weighted system. (Brandt *et al.* 1992, 272). Both environmental and social factors were included in the predictive model through this method. The weighted analysis of the criteria occurred in two phases, the first phase included the weighing of three types of proximities (0-1km, 1-5km, over 5km) from Roman water sources and roads. The assigned weight of the closest Roman road proximity (0-1km) was consistently higher than the weight of the river proximities. These weights were then combined at the end of this first phase. The second phase included the reclassification of the aspect and slope values to only give a moderate amount of weight to areas which were southern-facing and had a slope degree of under 10. No amount of weight was given to areas that did not meet these requirements. These two phases of deductive modelling were combined.

4.2.2. Inductive methods

The inductive phase of model building did not utilise the weighted system, but rather aimed to integrate the influence of Roman town locations. Site density analysis was conducted for the modelling sample data using a heat map. This heat map was then used to identify two classes of Roman settlements, major or minor. Minor densities were given influence zones of 1000 meters around the point, while major densities were given influence zones of 2000 meters. These zones were added to the deductively-created model as areas of 'Very High' predictive values.

The use of site density analysis on the Roman Hertfordshire dataset poses potential issues with the outcome of this method. As site density analysis can only be conducted inductively, its reliance upon the known archaeological data allows for observation biases to impact the outcome. The way archaeological data is often collected is nearly always non-randomly (Van Leusen 2002, 76), and this allows biases to occur within the data. For example, both the method in which the data was collected (Wilcox 2014, 344) and the research questions which prompted the discoveries (Verhagen & Whitley 2012, 56) can simultaneously impact the types of sites that are discovered. However, environmental factors may also play a role in this bias, such as through differences in modern land-use and soil textures (Van Leusen 2002, 76). These extraneous factors impact the basis of an inductive predictive model from its inception, thereby causing predictive values to be based on unrepresentative data and led to unfounded patterns.

To avoid these issues associated with the inductive method within the predictive model of Roman Hertfordshire, a predominantly deductive method was used. Known site data was then used in the creation of the model either to ensure the continuous improvement and accuracy of the different model versions through the count of sites per area, or to inductively infer where Roman towns were located within Hertfordshire in the final stages of the model's creation. As all potential biases should be understood, an assessment is required for both the environmental data and the archaeological data used to create and inform the predictive model.

4.3. Environmental Data Assessment

The environmental data used to inform and build the Roman Hertfordshire predictive model may have included biases towards the modern landscape, making it less applicable to the landscape in Roman Britain. This is especially relevant within the soil textures model, which included poor resolution data and may have been influenced by fertilisers or soil displacement. Carbonate content and soil depth were included in the soil data which has modern applications. Similarly, the layer of groundwater data included information only relevant to modern contexts, such as the flow mechanics and principal aquifers within the UK. A simplification was created for the soil textures and groundwater, reclassifying the attributes into broader, basic categories that were more likely applicable within the Roman period as they were based off of the underlying geology. The river vector layer was taken from modern contexts, but edited to include only the major branches of the river. However, it is possible that the river was located in a different place during the Roman era, and therefore puts questionable doubt upon the proximities of areas to water. More research may be required to understand the extent of river movement since this period, within this region of East England.

4.3.1. Future suggestions

In addition to describing and judging the quality of the environmental data, other forms of more relevant data can be included. Such forms include LIDAR elevation models which remove the elevation of vegetation, geological borehole databases, historical maps and remote sensing images (Verhagen *et al.* 2009, 22). These additional sources of data can be used in combination with the traditional modern environmental sources, where available.

4.4. Archaeological Data Assessment

Variables of depth and preservation which affect archaeology can impact their representation in the archaeological dataset, and therefore lead to the creation of predictive models that only take into account parts of the archaeological reality. These variables can be assessed with modern land-use, soil and groundwater data to account for this observational bias in the resulting model.

A way to analyse, and possibly correct, these biases is through a masking layer which can be created with all of the known sites within an area (Verhagen *et al.* 2009, 22). The modern land-use, soil textures and groundwater of the areas researched can be compared to the areas which contained no found sites. If there is an over-representation of a certain type of area, the source of the bias within the data could be identified and addressed in this way. It may also be assumed that if the distribution of sites is similar to the area distribution of certain soils or groundwater, then those layers likely are not the cause of the observational bias. It should not be assumed that this result indicates the lack of an observational bias entirely as other factors, such as method of observation, may contribute to this.

In the case of the Roman Hertfordshire predictive model, a 500 meter buffer was made around all known Roman sites within the area (n=4358). This buffer layer was cut from the boundary extent of Hertfordshire and given a translucent symbology, allowing a view of the areas which had identified Roman sites and those which had not. Using the 500 meter buffer around each site, a total of 22.6% of the area was 'researched', while the large majority had not found any roman sites (tab. 5)

Table 5: Area and percentage of the 'researched' areas and 'non-researched' areas in Hertfordshire, with researched areas determined by a 500 meter buffer around all known Roman sites (n = 4358).

Researched area	Non-researched area
372.089 km²	**1271.669 km²**
22.6%	77.4%

The masking layer was then applied to the modern land-use, soil group and groundwater maps. The area of each value within each map was calculated for the areas researched and were compared to the areas not researched.

4.4.1. Soil types

The researched areas appeared to not be overtly affected by the soil types present in the area of Hertfordshire (fig. 31). A fairly equal distribution of Roman sites (fig. 32, appendix 19) were identified among each of the five soil groups, according to their percentage of the total area of researched and non-researched soil. The most Roman sites were found within loamy soils (66.4%), however the highest soil distribution within Hertfordshire also is loamy soils. This high frequency is represented in both the researched and non-researched area (68.6%). The silt, sand and clay group had similarly small proportions within the researched and non-researched areas.

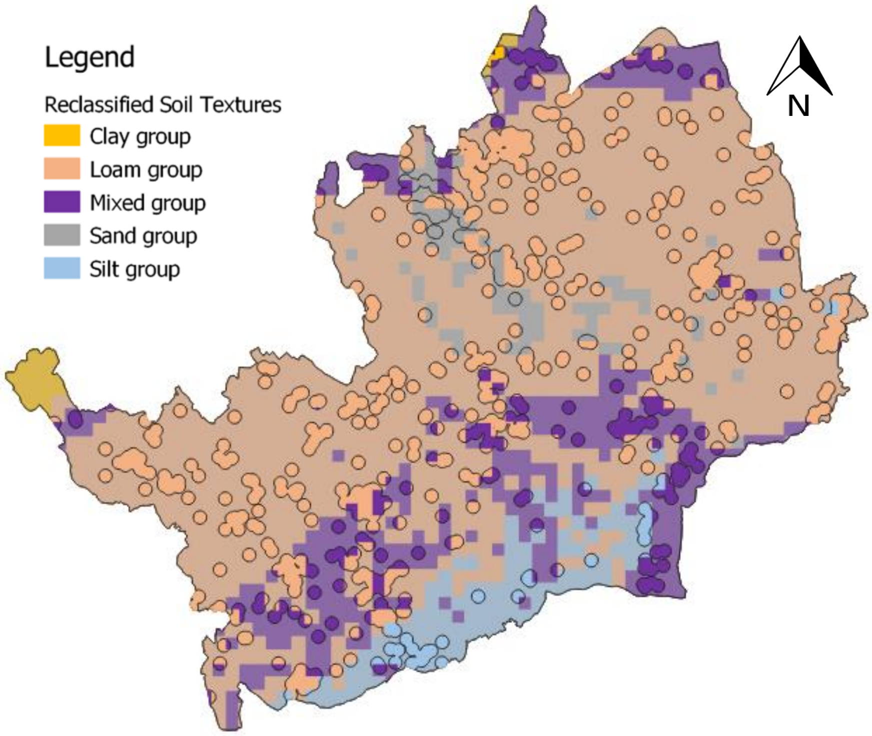


Figure 31: Distribution of non-researched areas in Hertfordshire (grey layer) and the underlying soil textures.

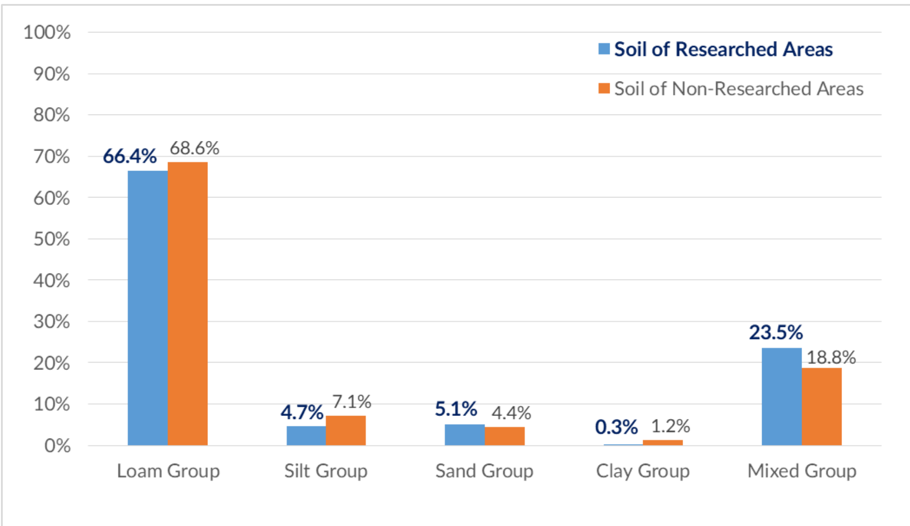


Figure 32: Frequency chart of researched and non-researched areas by their soil texture.

4.4.2. Groundwater

The groundwater areas fairly proportionate in the wet, damp and dry areas among both researched and non-researched areas (fig. 33). The similar levels of groundwater, with around 4% in difference between the two groups, suggests that groundwater may not have been a large bias within Roman site observations (fig. 34). This is perhaps the case because the high amounts of wet groundwater areas within Hertfordshire promote a generally good level of preservation of archaeology (appendix. 20). However, around 20% of area that was researched and not researched were 'dry', indicating less of a groundwater bias even on the basis of generally good preservation.

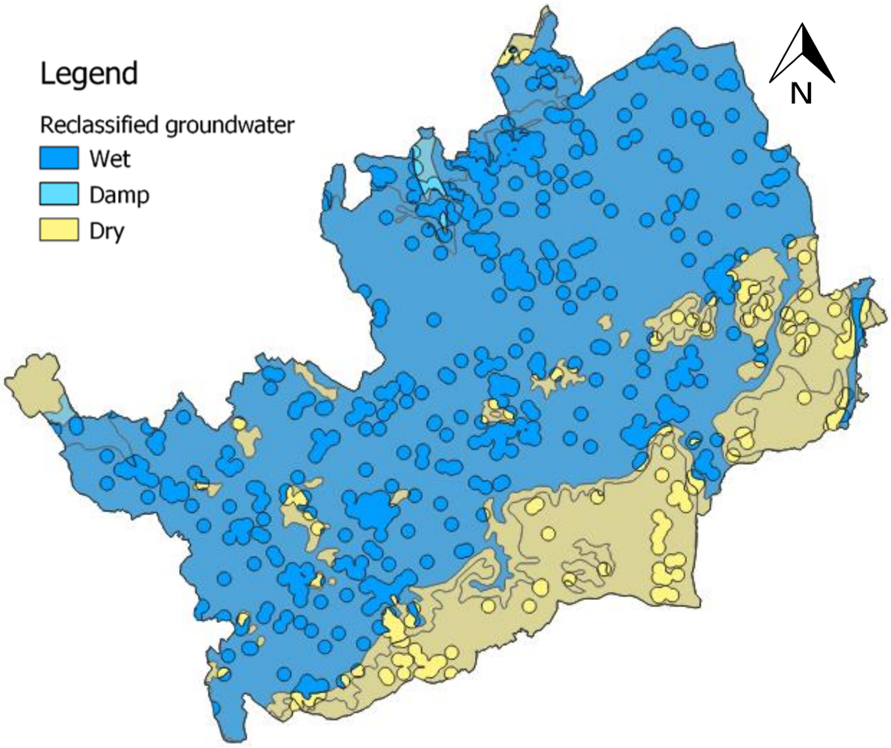


Figure 33: Distribution of non-researched areas in Hertfordshire (grey layer) and the underlying groundwater level.

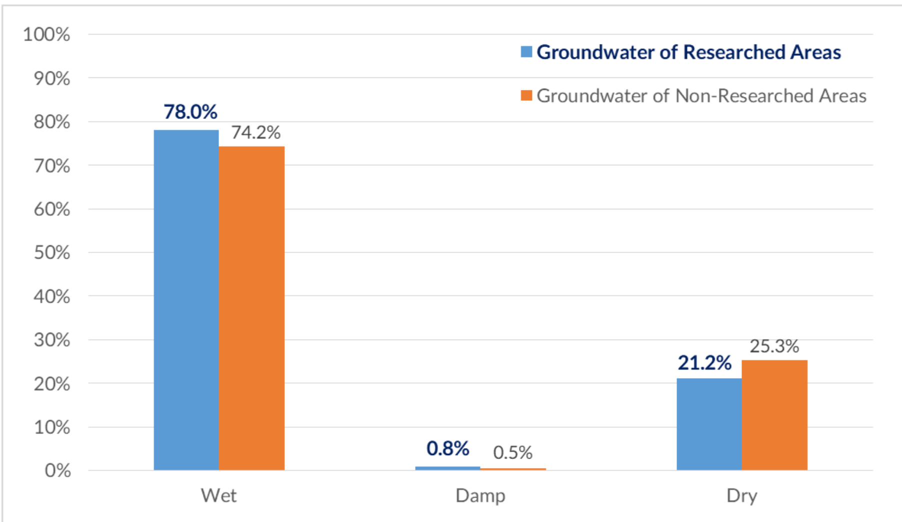


Figure 34: Frequency chart of researched and non-researched areas by their groundwater level.

4.4.3. Modern land-use

Within the layer of modern land-use, observational bias is likely present (fig. 35). This is due to the increase of development-driven archaeological projects after the creation of the Valetta Treaty. In regards to Roman archaeology in Hertfordshire, most sites were found within urban areas and croplands (appendix. 21). More of the urban areas were researched (35.4%) than not (17.4%), and the opposite is the case for areas of cropland (fig. 36).

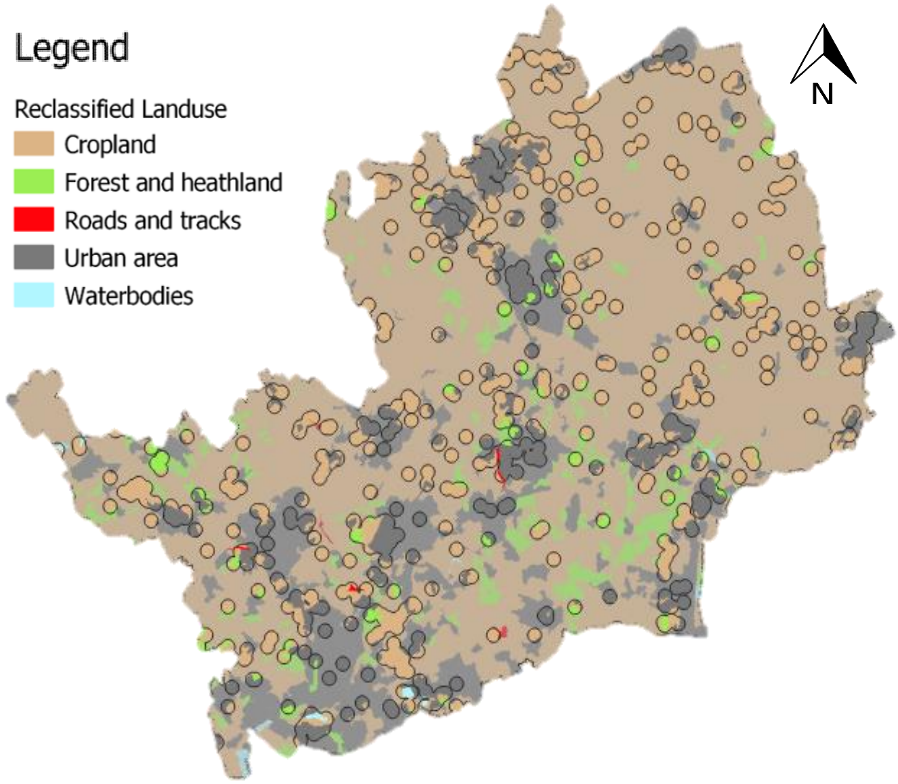


Figure 35: Distribution of non-researched areas in Hertfordshire (grey layer) among their modern land-uses.

The proportion of forest and heathland areas were fairly equal, which was also the case for the very few sites located in modern-day water bodies and areas of roads and tracks. Due to this unequal proportion of researched and non-researched urban areas, the archaeological data was likely biased towards development-driven archaeology. While not completely absent, research-motivated fieldwork has commonly been scarcer in modern town centres (Holbrook 2015, 2). Another potential reason for this proportion is the preservation of archaeology within cropland areas. Ploughing the land can cause disturbance of underlying archaeology, and therefore would make it more difficult to identify. However, surface-level archaeology may be visible from cropmarks visible in the open cropland.

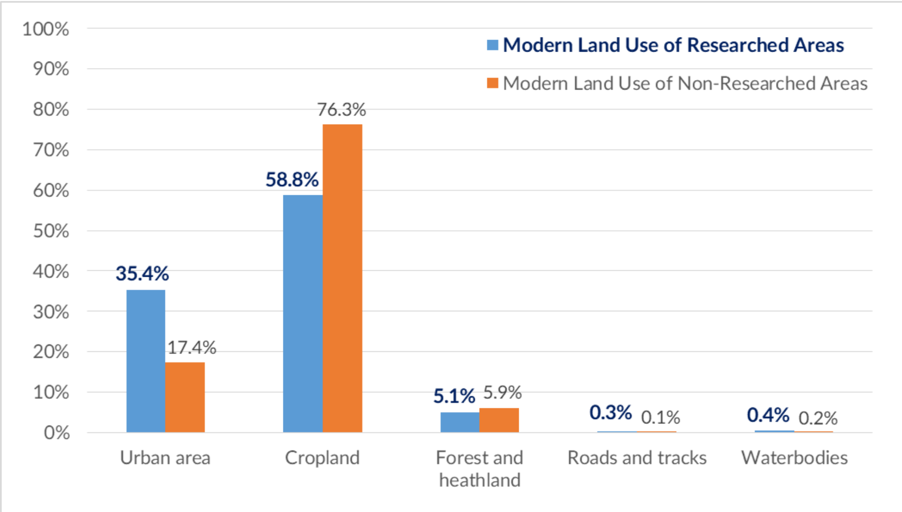


Figure 36: Frequency chart of researched and non-researched areas by their modern land-uses.

4.4.4. Future suggestions

Future recommendations for addressing the quality of archaeological input data include using sophisticated statistical methods to measure the weight of data biases, such as Bayesian statistics or Dempster-Shafer modelling (Verhagen *et al.* 2009, 22). Pre-emptive strategies can also be taken to improve the way of gathering and registering archaeological data in a national database (Verhagen *et al.* 2009, 22). Although, it must be taken into consideration that improving representability of data through methods of data-gathering may not be easy to implement in a country as large and geographically diverse as England.

5. Results

5.1. Application of the Methodology

The initial assumption was made that being closer to the road and river system would have produced positive side-effects in terms of habitability (Kamermans *et al.* 2004, 6). For most site types, but especially settlement and economic sites, water sources would have been needed within a reasonable distance (Brandt *et al.* 1992, 269). Therefore, the weighted system took this into account by valuing the close-distance zone from water with the highest weight, thereby increasing the area's archaeological predictability. The first step of applying the methodology stated previously was to individually explore the factors of water distance and road distance, and their presumed ability to predict the location of Roman aged sites in Hertfordshire.

5.1.1. Evaluating the proximities of rivers and roads

It was important to first evaluate the river and road factors individually before combining them into the model. This was done through the creation of two proximity rasters which were then evaluated with the site point data for their potential influence on site location.

As both the river and road layers were in the format of line-based vectors, the 'Rasterize' procedure was used to convert each layer into raster grid-based cells (pixels). This procedure created a new rasterized version of the Roman roads and river systems, with each cell being marked as either '0' indicating a blank cell, or '1' indicating the presence of water or road (fig. 37). With this new raster data, proximity raster layers were generated, also known as Euclidean distances (fig. 37). The proximity values were limited to 5000 meters away from the source of data, producing graduating values of distance for each cell within the layer extent, maxing out at 5000 meters in distance.

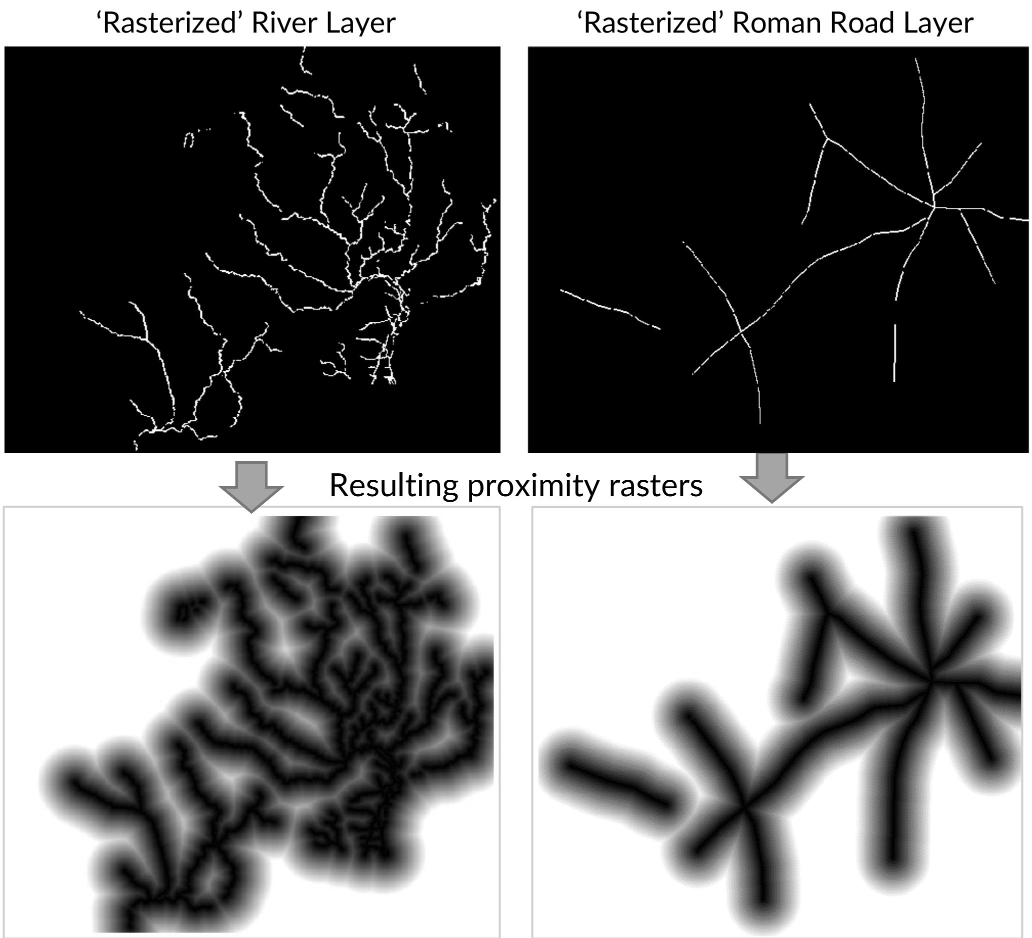


Figure 37: Results of the ‘Rasterized’ processing of the original vector files and the proximity rasters created from these rasterised layers.

The distance categories, weight, and overall importance of each factor’s proximity was then assigned through reclassification (tab. 6). The reclassing of distances was done using the QGIS ‘Raster Calculator’, through which an expression was written to separate and assign weights to three proximity groups (fig. 38). This process was applied to both the rivers and roads proximity rasters.

Table 6: Distance categories and weights applied to the river and road proximity rasters.

River Proximity		**Roman Road Proximity**	
Distance Categories	***Weight***	***Distance Categories***	***Weight***
0 - 1 km	**100**	0 - 1 km	**100**
1 - 5 km	**50**	1 - 5 km	**50**
>5 km	**10**	>5 km	**10**

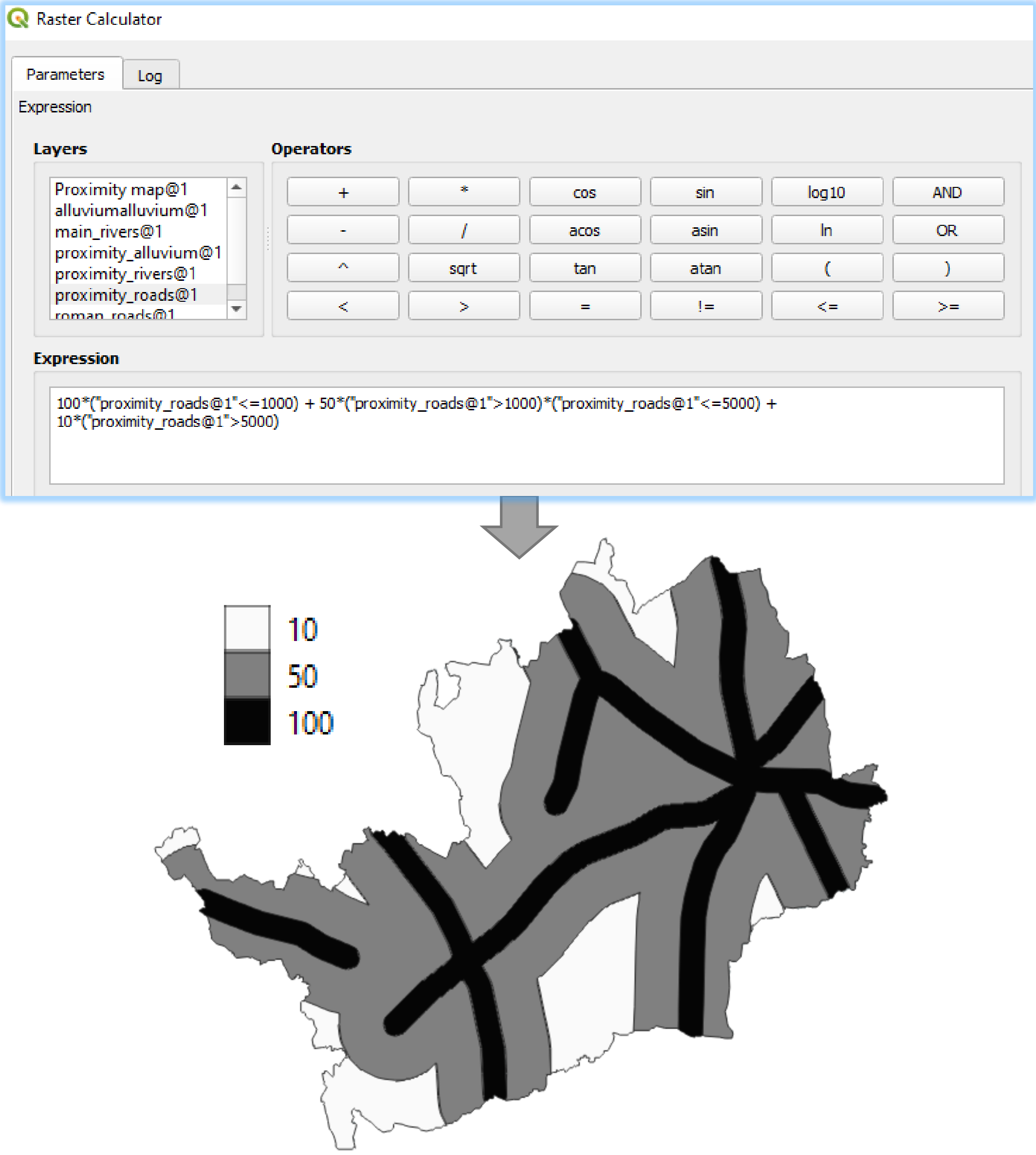


Figure 38: The reclassification query for the road proximity raster as shown in the QGIS 'Raster Calculator', and the resulting distance groups coloured by their weights.

The sample of known Roman archaeological sites (n = 3466) was layered on top of each of the reclassified proximity rasters in order to observe the count of sites within each of the three proximity categories (fig. 39 and fig. 40) in order to test its predictive value. The size (km^2) of each category was also calculated so as to take into account the category's proportion of the total area, in addition to the proportion of sites that are located in the area (tab. 7 and tab. 8).

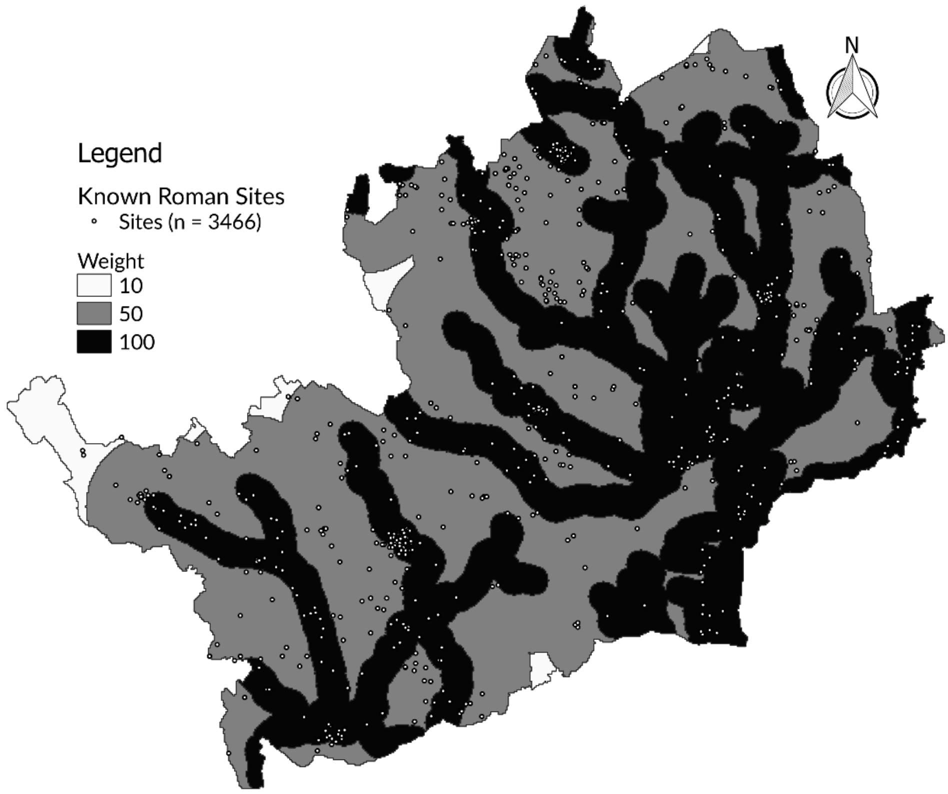


Figure 39: Reclassified river proximity raster with the layer of known Roman sites (n = 3466).

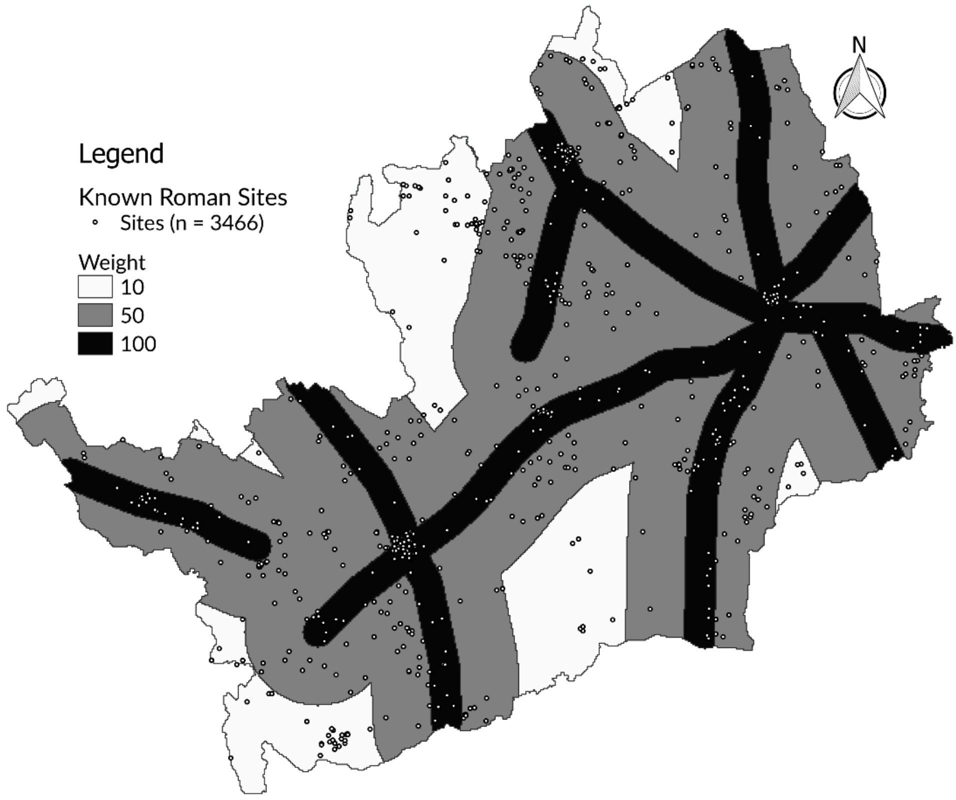


Figure 40: Reclassified road proximity raster with the layer of known Roman sites (n = 3466).

Table 7: The area (km²) and count of known sites in each proximity category from the river, expressed as the proportion of the total area (1643.7 km²) and site count (n = 3466).

River Proximity				
Distance	**Number of known sites**	**Site %**	**Area (km²)**	**Area %**
0 - 1 km	2313	**67%**	769.0	**47%**
1 - 5 km	1136	**33%**	834.0	**51%**
>5 km	17	**0%**	40.7	**2%**
Total:	3466 sites	**100%**	1643.7 km²	**100%**

Table 8: The area (km²) and count of known sites in each proximity category from the Roman road, expressed as the proportion of the total area (1643.7 km²) and site count (n = 3466).

Roman Road Proximity				
Distance	**Number of known sites**	**Site %**	**Area (km²)**	**Area %**
0 - 1 km	1691	**49%**	369.9	**23%**
1 - 5 km	1404	**41%**	971.8	**59%**
>5 km	371	**11%**	302.1	**18%**
Total:	3466 sites	**100%**	1643.7 km²	**100%**

Table 7 and 8 display the predictive potential of both of the proximity rasters for the rivers and Roman roads by using the known sites as a basis for site location preferences. In both of the proximity rasters, the 0-1km distance category locates a majority of the site sample, suggesting closer proximity to rivers and roads was favoured and influenced site location. When paying closer attention to the proportion of the areas of each distance category, the 0-1km area around the Roman roads appear to predict more sites (49%, tab. 8) within a smaller proportion of the total area (23%, tab. 8). The 0-1km area around the rivers locate more sites (67%, tab. 7) than the road proximity area, but does so within a larger area that constitutes almost half of the total research area (49%, tab. 7).

In addition to this, the nature of the river and road factors differ within the landscape and thus their influences in site location would also differ. The rivers signify an established environmental influence, which may have in the past been socially-influential through water transportation. Meanwhile, the Roman roads signify a newer, social influence in the landscape – one which may have also altered the influence of the river systems in the landscape.

Therefore, due to a higher surface area coverage by the river layer and the differing nature of the two factors, a direct comparison of their predictive potential is difficult to make. However, it can be stated through this evaluation that both layers can be combined to infer site location, and therefore possess key predictive capabilities.

5.1.2. Weighted proximity to roads and water sources

During this step of the modelling process, the two evaluated factors of water proximity and road proximity were to be combined, representing both a social and environmental factor in predicting Roman site location. Selected improvements were made to the proximity rasters before the layers were combined into the first prototype of the Roman Hertfordshire predictive model. One improvement made included an improved water proximity raster which takes into account water access from the main river system as well as other Roman water sources. As well as this, both proximity rasters were reclassified again to assign altered weights to each distance category.

Within the dataset, the site category of 'water sources' was created by grouping the records of water-associated structures which were identified as dating to the Roman era (fig. 41). The group contains archaeological 'subjects' such as wells and brick culverts (appendix 16). By merging the layers of various Roman water sources and the river system, the representation of which areas were in close proximity to water became more specific to the Roman era by taking into account the man-made infrastructure which enabled wider access to water.

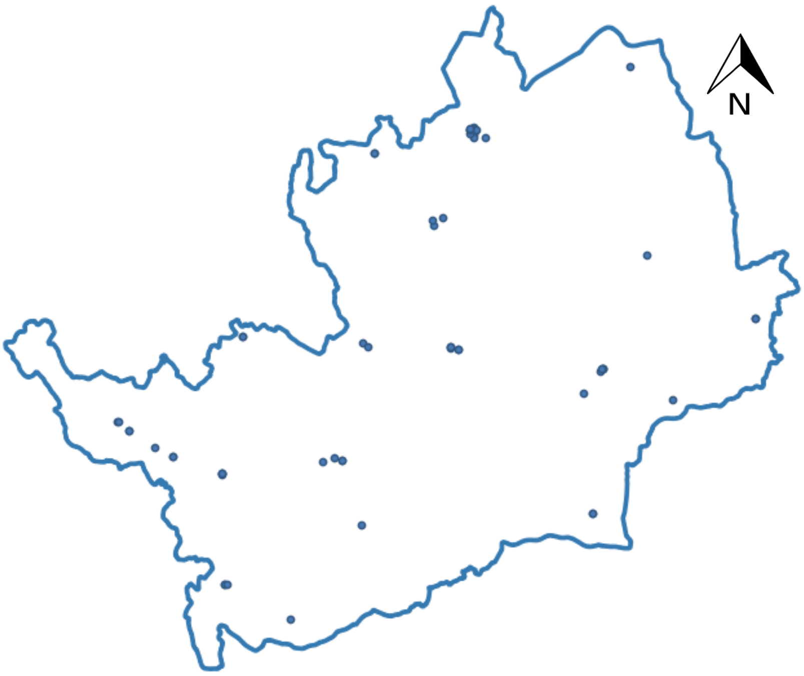


Figure 41: Known sites of water sources, dating to the Roman period (n = 62).

The previously used 'rasterized' river layer was merged in QGIS with the 'rasterized' water sources point layer that displayed the locations of the Roman water sources. This was done using the QGIS Raster Calculator to create a single raster layer with both instances of water access. The merged layer was then processed into a proximity raster with a maximum distance of 5000 meters, using the same process as used in the making of the original river proximity raster. The result included moderate yet sufficient differences from the previous water proximity raster (fig. 42).

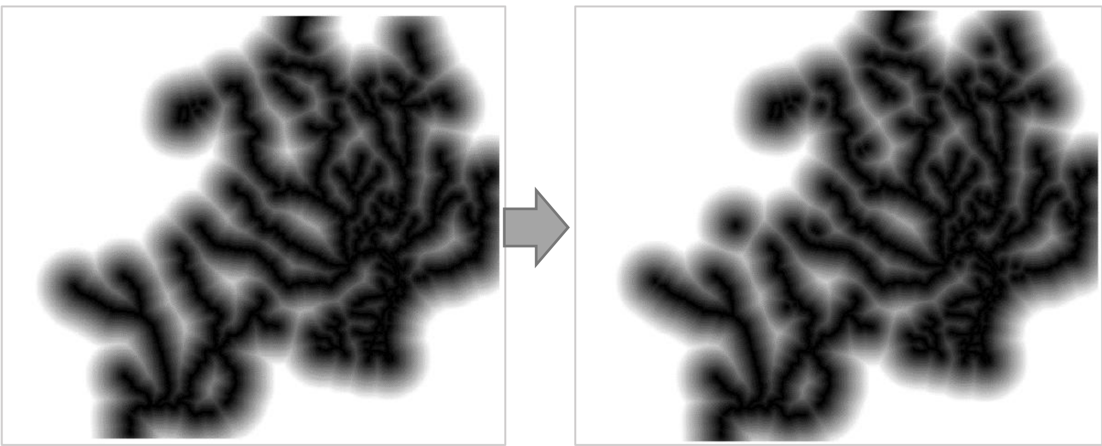


Figure 42: Improved proximity raster of water sources which includes both the main rivers and identified Roman water sources.

It appeared from the previous evaluation that, within Roman Hertfordshire, the flooding of land nearby to a river did not appear to be a concern in regards to site location. This was likely due to the drainage techniques practised by the Romans, marking the “first evidence of extensive drainage in Britain” (Brown 1997, 269). Therefore, the weight of the closest proximities to water remained the highest.

However, it was decided that the proximity of the road systems should hold greater weight in the multi-criteria analysis because of the connections roads came to provide to key centers in the Roman era. This could have provided an incentive for close proximity to either major or minor roads for both residential and commercial sites. For example, ‘Watling Street’ passes through the major Roman town, Verulamium, in Hertfordshire and travels south-eastwards to the major commercial center of Londinium (fig. 43), while ‘Stane Street’ was said to have linked centers like Verulamium to Colchester (fig. 43) (Fulford 2015, 75), the largest center in Roman Britain. This social factor was likely to have impacted the locations of economic and settlement sites, and perhaps ritual or burial site locations also.

Figure 43: Roman roads which passed through the area of Hertfordshire.

Due to these theoretical viewpoints, the reassignment of weights for each distance category was undertaken for both the water and road proximity raster layers (tab 9). The result of these new weights and their visualisation in QGIS are displayed in Figure 44.

Table 9: Adjusted distance categories and weights applied to the water and road proximity rasters.

Water Proximity		Roman Road Proximity	
Distance Categories	***Weight***	***Distance Categories***	***Weight***
0 - 1 km	**80**	0 - 1 km	**120**
1 - 5 km	**50**	1 - 5 km	**90**
>5 km	**10**	>5 km	**30**

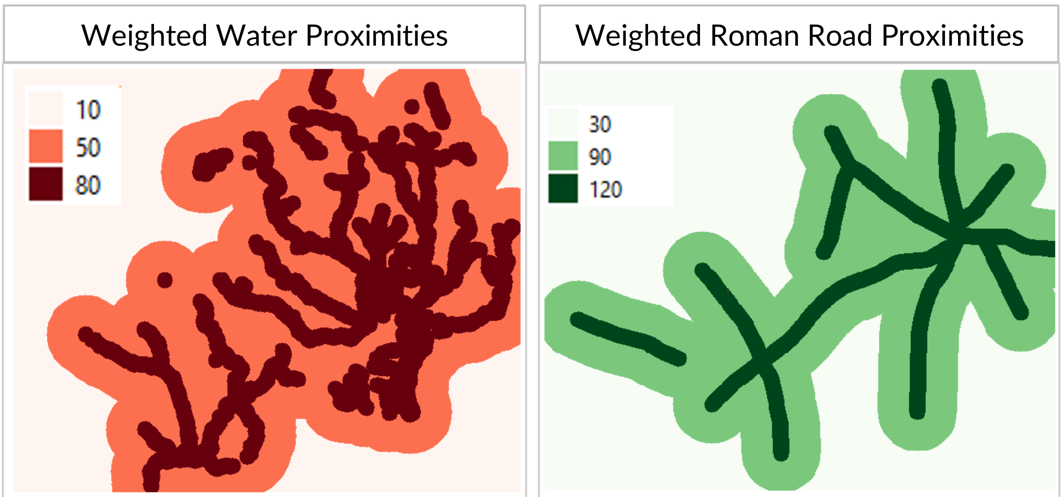


Figure 44: Reclassified weights of the water and Roman road proximities.

Both of the newly reclassified layers were used in the multi-criteria analysis through the merging of the weighted cells in the QGIS ‘Raster Calculator’. The result was multiplied by the raster boundary to discard cell values outside of the extent of Hertfordshire. This produced the first version of the Roman Hertfordshire predictive model, ‘Model 1’ (fig. 45), in which cells within 0-1km proximity of the roads and rivers had overlapped to produce the highest predictive value areas (tab. 10).

Table 10: Total weight of each predictive value in Model 1.

Model 1	
Weight	**Predictive value**
200	Very High
160	High
120	Medium
80	Low
40	Very Low

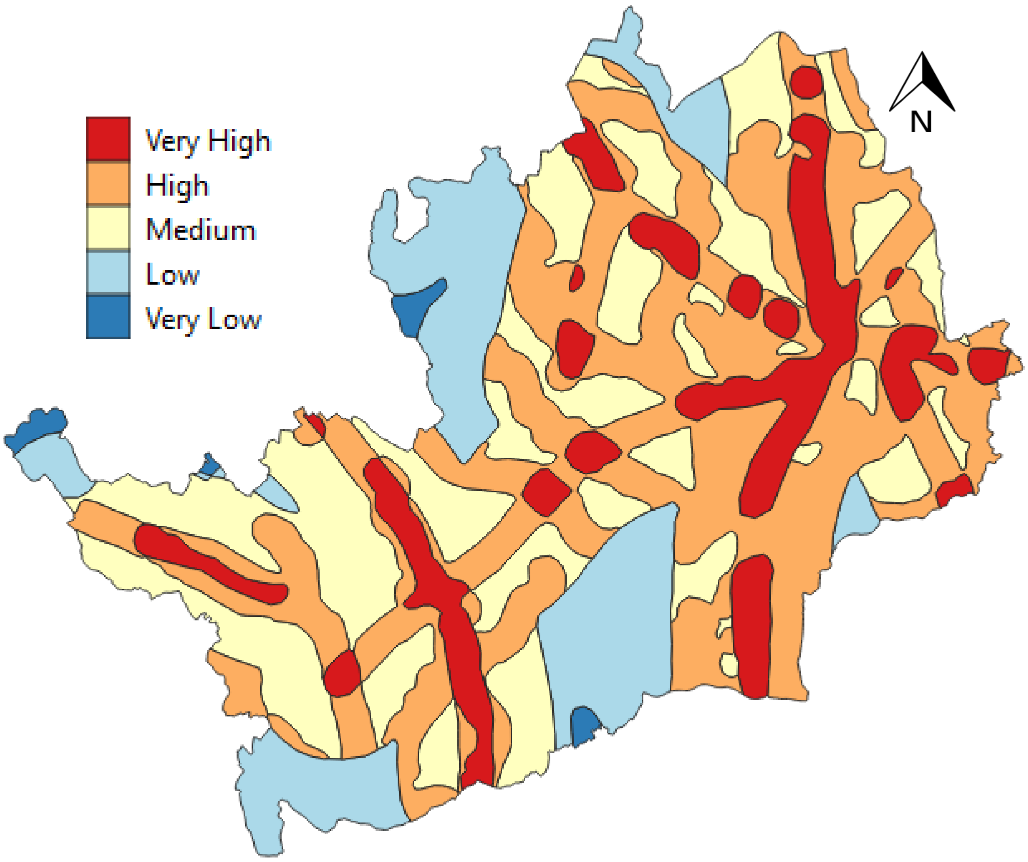


Figure 45: First version of the Roman Hertfordshire predictive model (Model 1), created by multi-criteria analysis of proximity to water and Roman roads.

The resulting raster model was digitised into separate vector polygons while using the raster model as a basis. Each predictive category was merged into a single attribute in order to calculate the total count of known sites within each category (tab. 11). The known site count within the 'Very High' and 'High' areas totalled 72.4% of the site sample, whilst covering around half of the area of Hertfordshire (53%, tab. 11).However, the 'Very High' category constituted only 14.1% of the total area but counted 45.2% of the known sites within this small percentage. This is an important point to note, as a predictive model should not only aim to be correct, but also precise in its predictions. If a model predicts over half of a given area is 'Very High' in archaeological prediction, it creates difficulty in the extent of research that can be done over such a large area. Therefore, it is necessary to attempt to narrow down as far as possible areas which show the highest level of archaeological value. The 'Very Low' predictive category contained no site occurrences, however it only constitutes 1% of the total area, making it clearly the smallest category.

Table 11: The total count of sites (n = 3466) and area within each predictive category in Model 1.

Predictive value	Number of known sites	Site %	Area (km²)	Area %
Very High	1566	**45.2%**	232.141	**14.1%**
High	944	**27.2%**	638.949	**38.9%**
Medium	598	**17.3%**	466.905	**28.4%**
Low	358	**10.3%**	289.768	**17.6%**
Very Low	0	**0.0%**	15.999	**1.0%**
Total:	3466 sites	**100%**	1643.758 km²	**100%**

With that being said, additional site location factors should be integrated into the multi-criteria analysis in order to take into account alternative influences. For example, while the distance from roads would be an important factor for all site types, elevation-derived layers also have the potential to predict which areas were optimal in landscape for site location.

5.1.3. Weighted aspect and slope

The second version of the Roman predictive model of Hertfordshire builds upon the first model through the integration of the derived layers of slope and aspect in Hertfordshire. The integration of slope and aspect would benefit the ranking of areas for Roman site prediction by their ability to identify optimal areas for construction. The theoretical assumptions used to identify these optimal areas included both the degree of slope at, or below, 10 degrees and a southern-facing aspect degree. In areas where the slope was below 2 degrees, the aspect could be facing any direction as the land would be flat enough to not limit solar radiation.

In the northern hemisphere, where England is located, the northern side of slopes would often be shaded and would receive drastically less solar radiation. The importance of solar radiation that is successfully received from the sun is “the primary energy source that drives many of the earth's physical and biological processes”, and therefore would be important when deciding the location of sites (www.pro.arcgis.com). This factor is especially the case for agricultural sites since abundant sunlight (in addition to water access) is needed for the cultivation of crops. However, these conditions would also have been optimal for construction of structures, whether residential or commercial in nature, as well as for animal husbandry. These conditions may have not been sought after for ritual-related or military sites, as often high elevations and line-of-site were associated with such structures (Verhagen *et al.* 2007, 206), or ease of access (Wilcox 2014, 341). To represent this theoretical assumption in my predictive model, both the slope and aspect raster layers were re-classed using the QGIS ‘r.reclass’ tool.

Firstly, the aspect layer had one type of reclassification which aimed to only select parts of the landscape that were south-facing. To do this, a rule file was written within Notepad which classified that all cells with values between 0 and 111 should equal ‘0’, values between 112 and 247 should equal ‘1’, and finally that values between 248 and 360 should also equal ‘0’. The rule system was created on the knowledge that during the aspect layer’s creation, each raster cell was assigned a value that reflects its direction on a 360 degree axis. In order to only include the southern-facing areas, the diagram seen in Figure 46 was used to include all south-east, south and south-west-facing cells (112-247 degrees). Within the ‘r.reclass’ tool, the aspect layer and the rule file was selected as the terms for the layer’s reclassification. This procedure resulted in the layer seen in Figure 47, in which the white cells have an aspect value that is south-facing.

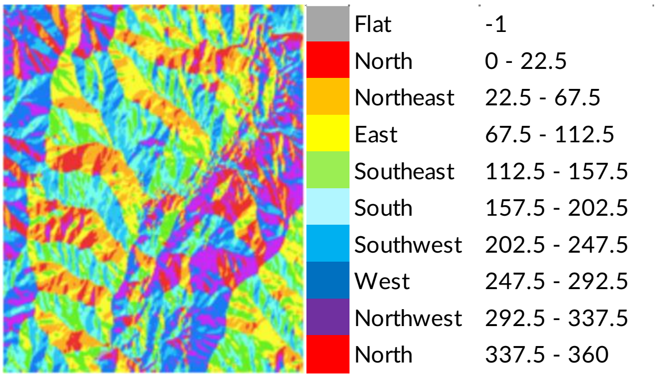


Figure 46: Diagram illustrating how the degrees of aspect determine the cardinal direction of a hill or mountain face in the Northern hemisphere (left image: www.pro.arcgis.com).

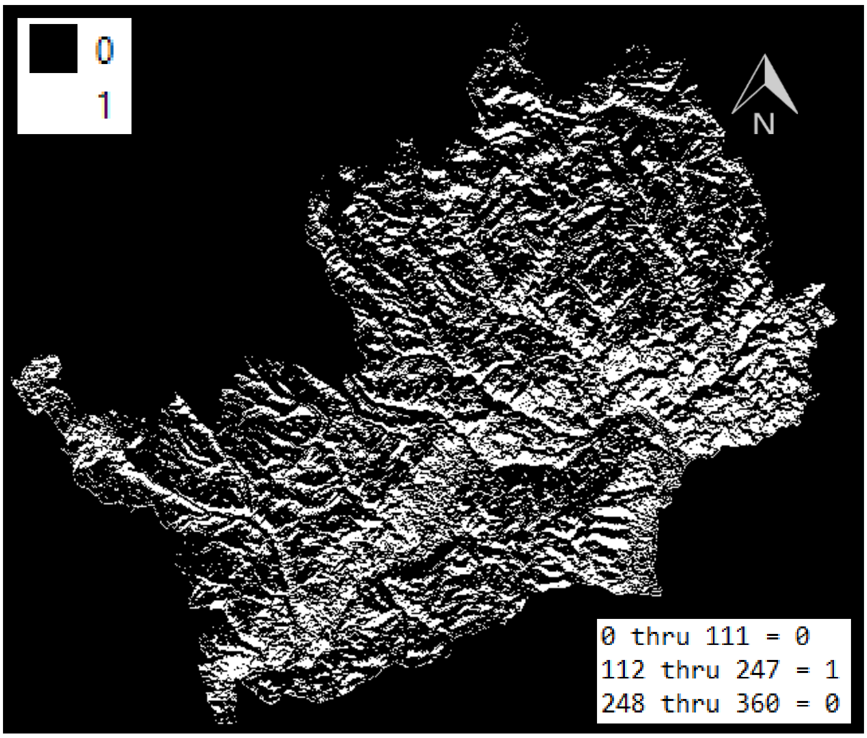


Figure 47: Southern-facing aspects and the rule used to define the reclassification.

The slope raster layer was used to make two separate reclassified layers with the ‘r.reclass’ tool: one in which the value of ‘1’ is given to cells where the slope value is 2 degrees or less (fig. 48), and another where the value of ‘1’ is given to cells where the slope value is 10 degrees or less (fig. 49). For this, two rule files were written which stated each condition, with those cells not fitting the requirements being classified as ‘0’.

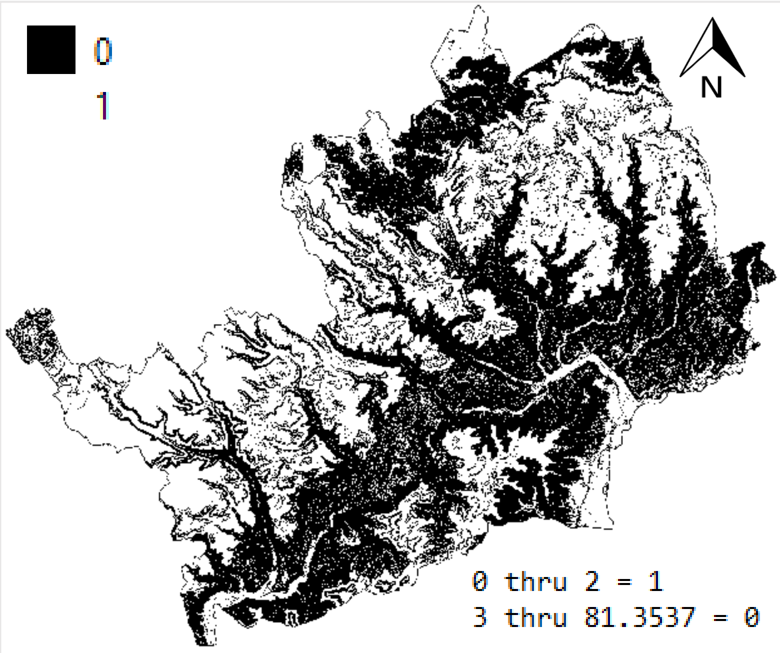


Figure 48: Slopes of 2 degrees or less, and the rule used to define the reclassification.

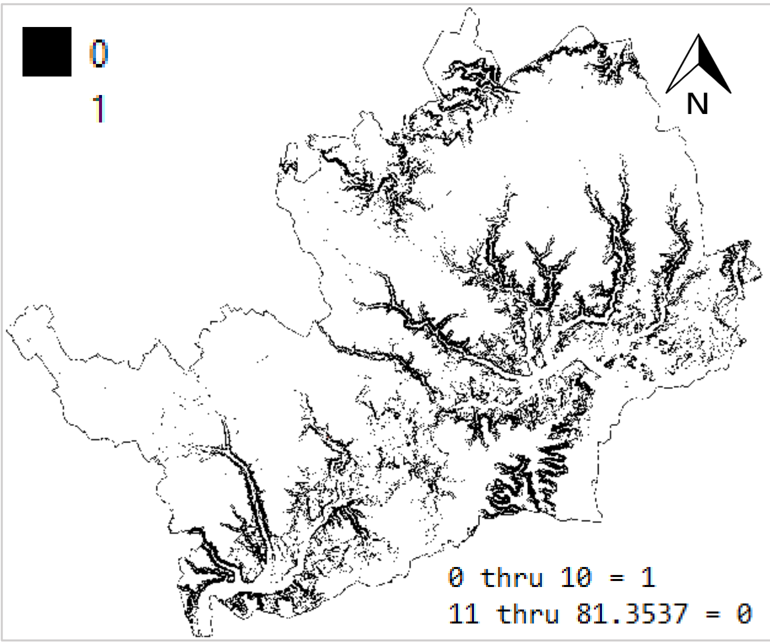


Figure 49: Slopes of 10 degrees or less, and the rule used to define the reclassification.

Now that both of the slope layers and the aspect layer were reclassified, they were to be merged into one layer with the conditions that cells should have a value of '1' if they are south-facing and their slope value is less than 10 degrees, or if their slope value is less than 2 degrees. This was expressed in the QGIS 'Raster Calculator' as:

```
( "aspect_south@1" = 1 AND "slope_less_10@1" = 1 ) OR
"slope_less_2@1" = 1
```

The expression written above combined these layers and produced Figure 50, which displays the values that met either of the two conditions as a white cell with the value of '1'. After this process, the assigned weight can be added to both the factors of slope and aspect, thus representing the influence of the terrain on Roman site location.

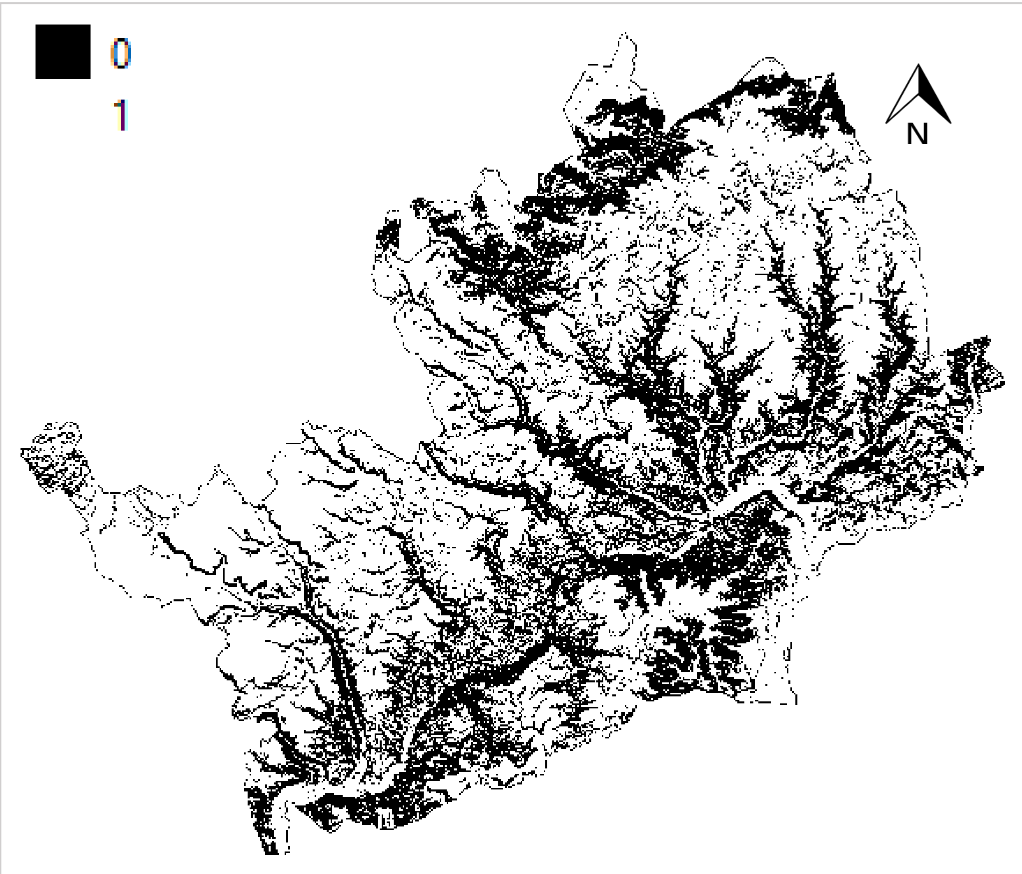


Figure 50: Model of optimal slope and aspect, with white cells not exceeding 10 degrees in slope and having a southern-facing aspect when solar radiation may be limited by the placement of the slope.

In order to combine this layer of optimal slope and aspect to the result of the previous predictive model version, a weight must be assigned to all cells with the value of ‘1’. Deciding what the weighted value would be of the layer was a difficult task as the relevance of aspect and slope is highly dependent on the site type. However, for this model a generalised weight of ‘30’ was chosen. This was decided because when the new weighted layer is merged with Model 1, the score is only able to promote a cell’s value to the next predictive group if it is in the higher tier of the previous group it was in. This would allow the influence of the terrain’s slope and aspect to influence the weighted system while not skewing the overall result.

The weight was assigned using an expression in the QGIS ‘Raster Calculator’, adding the aspect and slope weighted cells to the values of the previous predictive model layer, containing reclassified proximities to Roman roads and water sources. The result was also multiplied by the boundary raster so as to remove out-of-bounds pixel values:

```
( "deductive_predictive_model2@1" + "weight_30_as@1" ) * "boundary@1"
```

The output of this multi-criteria analysis (fig. 51) appeared to need visual simplification for its use as a predictive model, as the model’s representation of predictive values should aim to remain a generalisation of the archaeological situation. In addition to this, as the number of unique weighted values within the raster cells increased, categorising of each value into a predictive group was needed. In the final product, the influence of the slope and aspect of the terrain should not be directly noticeable, as it was decided more weight would be put on the proximity of areas to water sources and Roman roads.

Once the raster output layer was categorised as five main predictive values, it was decided against manually digitising the layer in order to produce the model in vector format. Instead, multiple QGIS and GRASS (Geographic Resources Analysis Support System) procedures were used to create the vector layer and generalise the result.

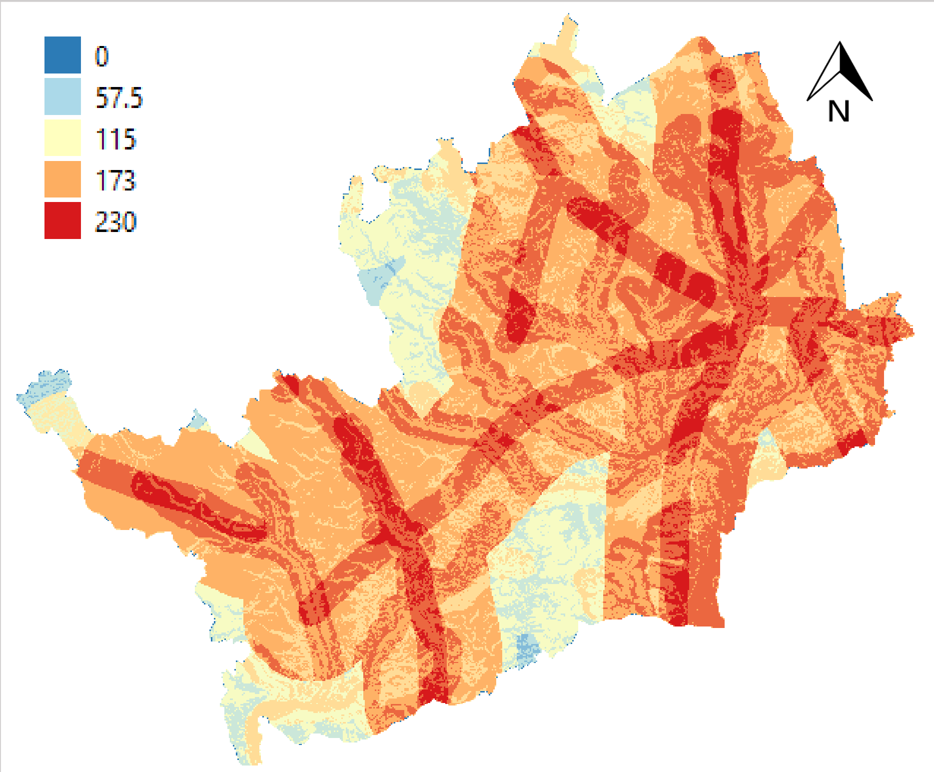


Figure 51: Uncategorised raster output layer of the merging of the weighted proximities from water and Roman roads with the weighted optimal slope and aspect raster.

The 'Polygonize' conversion tool was used to initially transform the raster pixels into vector polygons. A new attribute field was created for this vector layer, specifying the predictive value, ranging from 'Very High' to 'Very Low', and the symbology was categorised through this new field. This still left many separations in the polygons of each group, originating from the conversion of raster to vector. The separated polygons were fixed by editing separate selections of each group at a time and using the 'Merge Selected Features' digitising tool to create a single field for each of the predictive values (fig. 52).

Manual selections of certain areas was necessary to remove traces of pixels, such as from the 'Low' valued area in the west of Hertfordshire. From a wide-view, the model seemed to have smoother edges, however pixilation remained around the edges of each area. In order to reduce the pixilation caused by the original raster layer, the GRASS tool 'v.generalize' was used in conjunction with the "snakes" generalization algorithm selected with a maximum tolerance value of '1' (fig. 53).

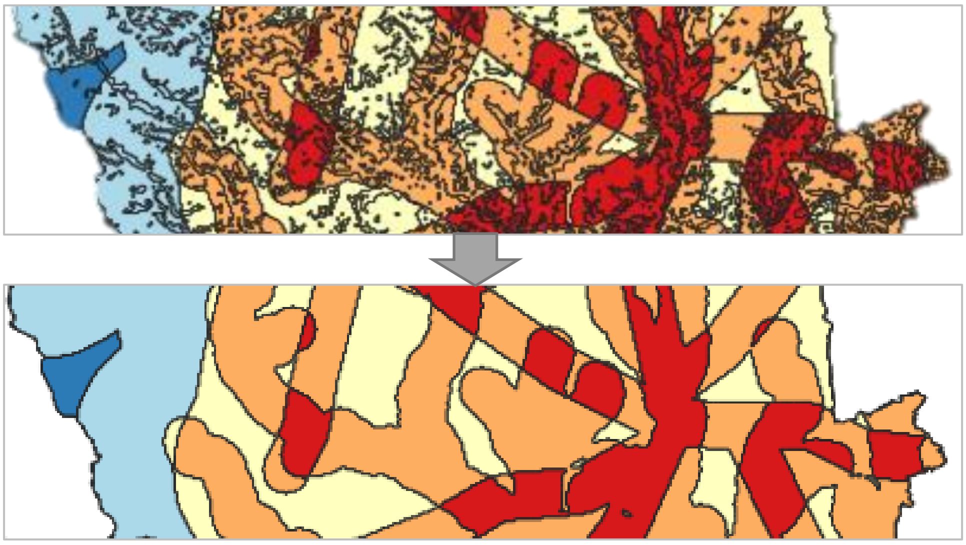


Figure 52: The before and after of the 'Polygonized' raster layer once the predictive value categories had been merged.

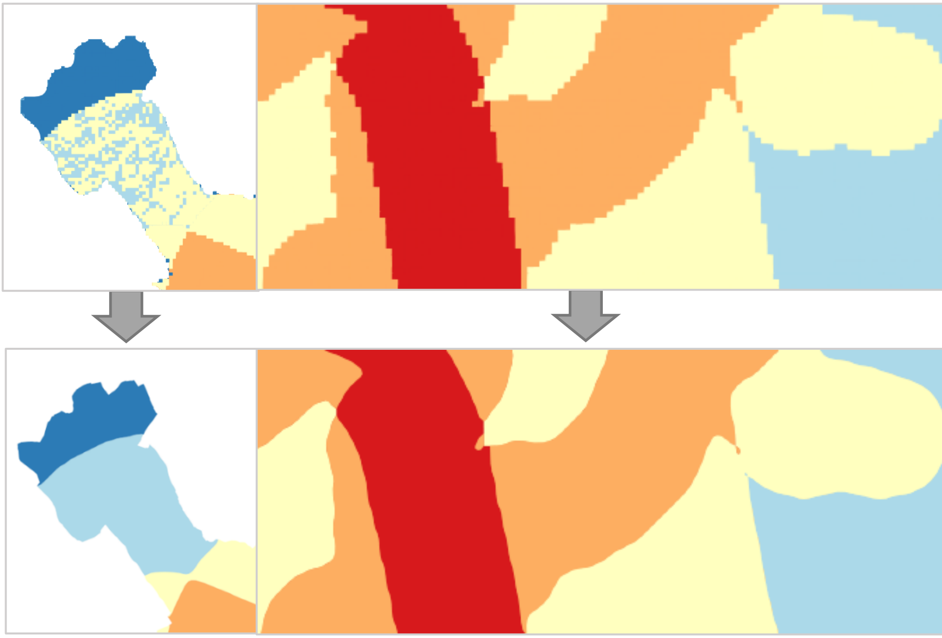


Figure 53: The before and after of the smoothing process, using the 'generalize' tool.

Through the various processes the polygon layer had been put through, the boundary edges of the model became fractured and was unable to fit the full extent of the Hertfordshire boundary. This was an issue for a few reasons, one being that the presentation of the model on close inspection is visibly worse. A second reason to fix this issue was that it likely would affect the count of known sites within the areas of the polygons, as well as impact the accuracy of the area calculations for each predictive group – which is needed for the Kvamme's Gain test of reliability and accuracy (Kvamme 1988, 329). In order to address this issue, the nodes of each bordering polygon were edited manually to extend the layer beyond the Hertfordshire boundary. During this process, small pixels that were found also were removed and blended into the surrounding predictive values. The extended polygons were then trimmed to match the extent of the county using the 'Clip' tool (fig. 54).

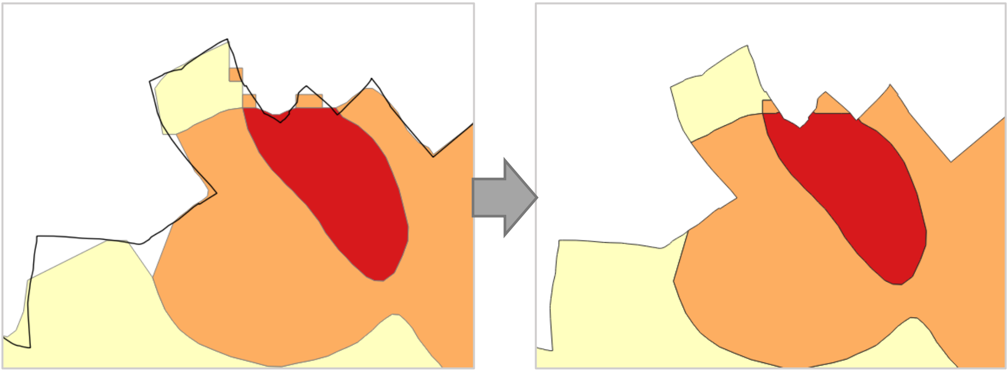


Figure 54: Before and after of the editing of the borders of the predictive modelling, using the 'clip' tool.

This version of the predictive model added more nuanced areas of medium, low and very low predictive values, especially in the central outskirts of Hertfordshire. More areas which were previously classified as 'Low' are now in the 'Medium' category (fig. 55) which is shown by the increased area percentage of the 'Medium' category from 28% in model 2, to 35% in this version (tab. 12). The 'Low' category naturally decreased in area by around 8%.

The impact these area changes had on the known site count can be seen especially in the 'Medium' and 'Low' categories. The 'Medium' category increased in site count by around 6%, while the 'Low' category decreased both in size and site count by 5% and 6% respectively. It would be beneficial for the model's precision to locate a higher percentage of sites within the 'Very High' category, while not increasing the area percentage substantially. However, it must be kept in mind that any change in predictive values should be based in theory that does not simply derive itself from the collected dataset as it may or may not be representative of all Roman sites.

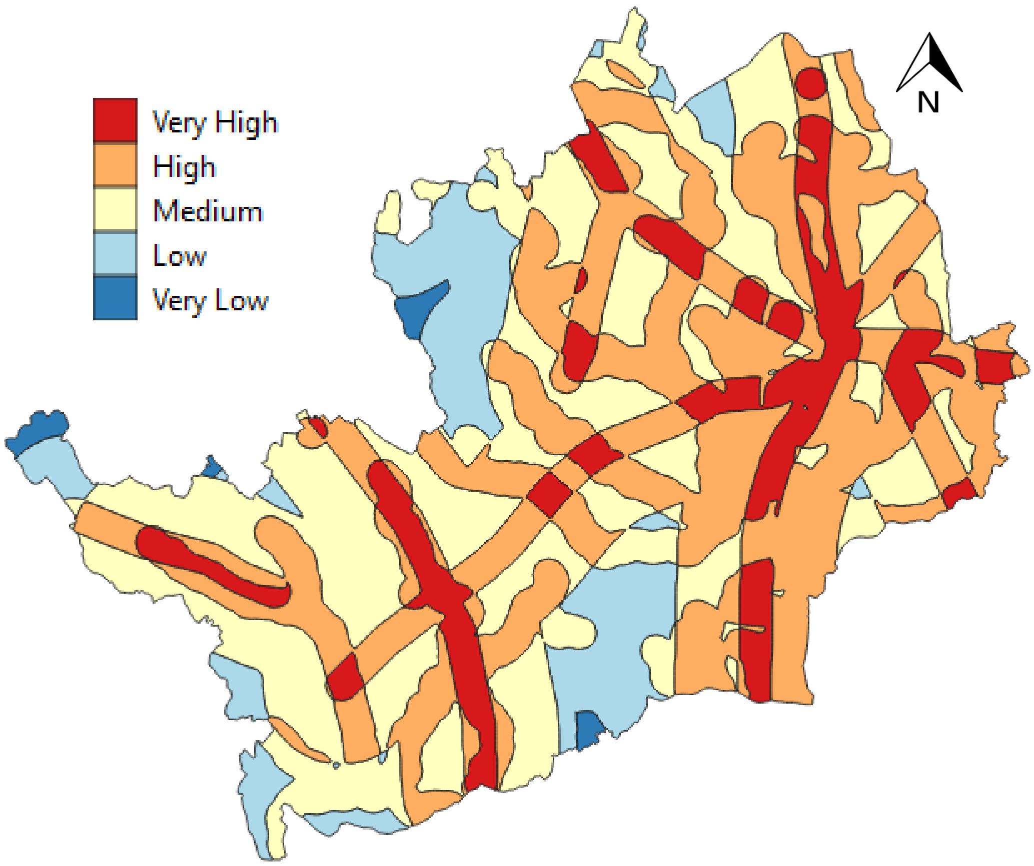


Figure 55: Second version of the Roman Hertfordshire predictive model (Model 2), created by multi-criteria analysis of proximity to water, proximity to Roman roads and optimal slope and aspect.

Table 12: The total count of sites (n = 3466) and area within each predictive category in Model 2.

Predictive value	Number of known sites	Site %	Area (km^2)	Area %
Very High	1533	**44.2%**	210.939	**12.8%**
High	951	**27.4%**	640.730	**39.0%**
Medium	815	**23.5%**	582.021	**35.4%**
Low	167	**4.8%**	194.555	**11.8%**
Very Low	0	**0.0%**	15.517	**0.9%**
Total:	3466 sites	**100%**	1643.758 km^2	**100%**

5.1.4. Site densities and Roman towns

Thus far, the Roman site observations had not yet been explicitly used in the theory behind the Hertfordshire predictive model. This ensures that unrelated factors to site location preferences, such as modern-day observation bias, does not impact the predictive model to a large extent. However, due to the large amount of known sites within Hertfordshire that have been identified as Roman, the density of the sites can be used to infer where Roman towns may have been located.

Through this knowledge, an area within these zones can be seen as very high in archaeological predictive value. The presence of Roman towns represents a socially-driven, and perhaps economically-driven, influence on site locations (Brandt *et al.* 1992, 269). For some sites, a location in close proximity to a center would have been intentional, and therefore potentially able to predict. To model this factor, heat map symbology was used on the point layer containing the 80% sample of known Roman sites (fig. 56).

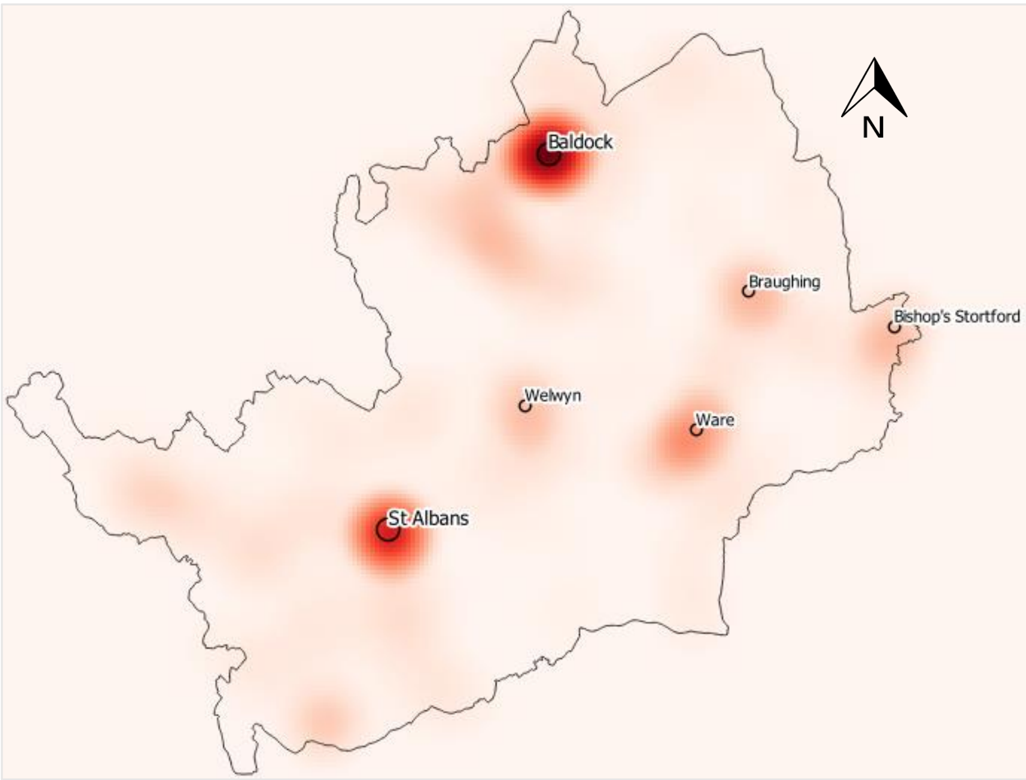


Figure 56: Heat map symbology that was used on of known Roman sites (n = 3466) with modern name labels on the most distinguishable heat spots.

The heat map symbology uses density analysis to determine where the largest concentrations of points are located and displays this through the use of ‘heat spots’. The heat spots were used as a reference to make two series of points in a new layer: one being major towns that are indicated by the most vivid heat spots, and the second being the minor towns that are indicated by a less defined heat spot. The Roman site data field called “Civil Parish” contained the modern town names where each observation was found, this field was used to name each major and minor heat point. The network of roads which were imposed after Roman invasion linked smaller developing urban and commercial centers, including Welwyn, Braughing, Ware and Baldock (Tereszczuk 2004, 10). There is little written about the developing centers which are identified as minor Roman centers on the heat map, therefore further explanation will only be given for the two major Roman centers located in St. Albans and Baldock.

Verulamium (modern-day St. Albans) was the third largest town in the Roman province of Britannia (Lockyear & Shlasko 2017, 17) and was located on the river Ver within Hertfordshire (Fulford 2015, 61). The center had direct links to other large Roman centers, such as Londinium (Roman-era London) and Colchester (Fulford 2015, 75), through the road system. The town steadily grew after its Roman invasion, with one of the earliest stone buildings appearing in Verulamium being a forum-basilica (Lockyear & Shlasko 2017, 19). A large town wall surrounded the center, with Roman bath houses and a theatre located within its boundary (Lockyear & Shlasko 2017, 19).

Baldock was another possibly large town within Roman Hertfordshire, also becoming one of the largest settlements in Roman Britannia (North Hertfordshire Museum 2019, 6). Extensive use of the Roman roads that pass through Baldock was evidenced by the layer of soil build-up identified over the road material, as well as the secondary fills of the roadside ditches (Phillips *et al.* 2009, 94). A series of boundary ditches marked the extent of the settlement (Phillips *et al.* 2009, 89).

The two major heat points were given a 2000 meter round buffer, using the vector 'Buffer' tool. All minor heat points were given a 1000 meter round buffer, setting its sphere of influence to be half as far. The two point layers were then merged into one and then subtracted from the last instance of the predictive model by using the vector geometry process 'Difference'. The output of that process was then merged with the point layer again to create no overlapping geometry. The buffered points were then selected, along with the other polygons in the 'Very High' predictive value category and merged using the 'Merge Selected Features' tool (fig. 57).

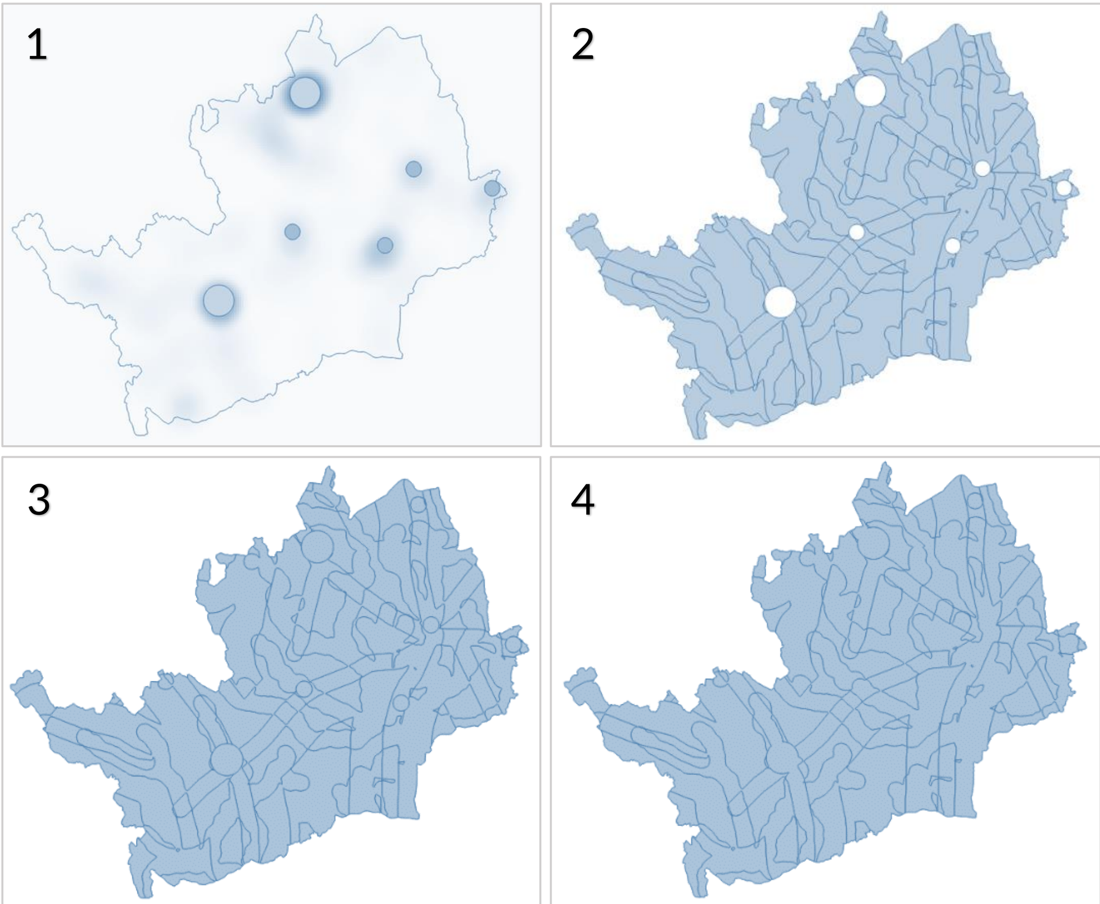


Figure 57: The four-step process to include the influence of Roman towns in the model. (1) Adding buffers around major towns (2000m) and minor towns (1000m). (2) Using the 'difference' tool to remove these areas from Model 2. (3) Adding the buffers back into the model. (4) Merging the buffers to the 'Very High' predictive category.

The addition of the buffered points to the final model slightly increased the area of the 'Very High' predictive category (13.6%) in aim of including areas in close proximity to Roman towns (fig. 58). The other influencing factors which were integrated into the model previously include proximity to Roman roads and water sources, as well as certain aspect and slope degrees. The influences of other environmental factors, displayed in the layers of groundwater and soil textures, will be discussed within the results chapter. However as for the lack of inclusion of soil and groundwater in the predictive model, prior to the creation of the predictive model, both layers were looked at spatially in terms of possible influences of known site locations. As a large majority of the area of Hertfordshire had high levels of groundwater (75%) and loam-textured soils (68%), associations with site locations could not be made.

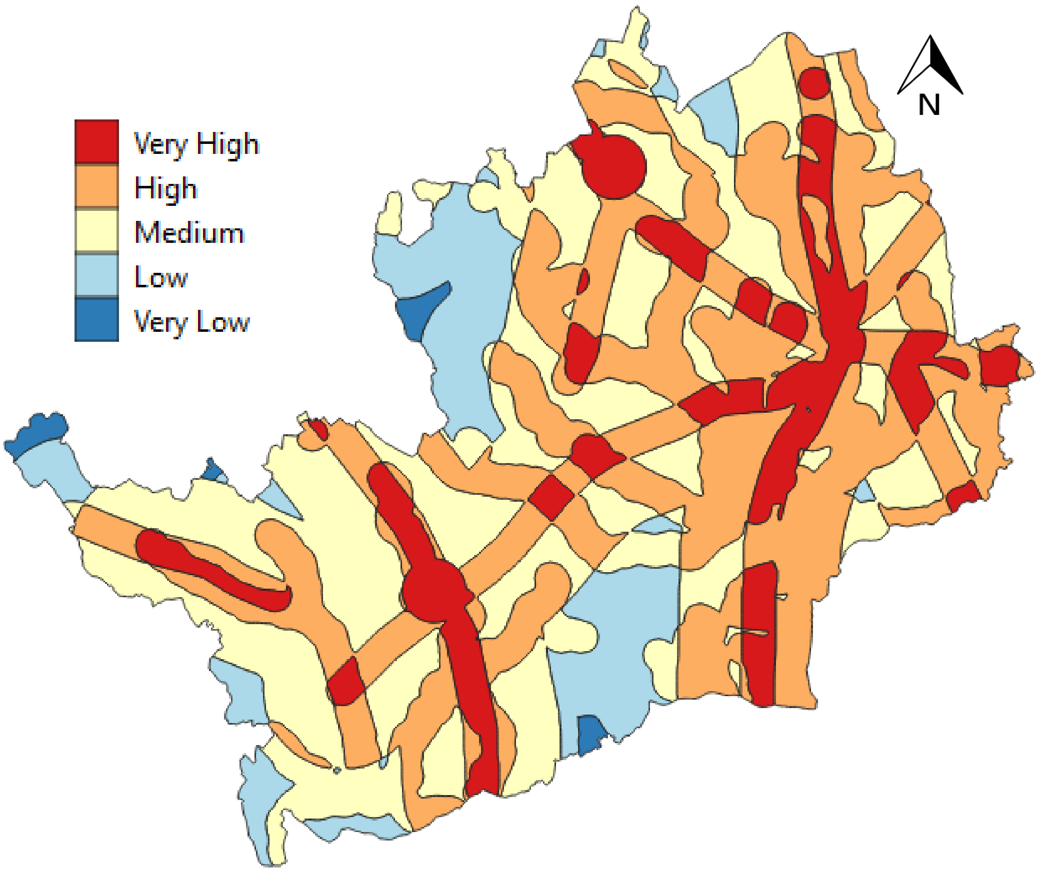


Figure 58: Third and final version of the Roman Hertfordshire predictive model (Model 3), created by multi-criteria analysis of proximity to water, proximity to Roman roads, optimal slope and aspect and close proximity to major and minor Roman towns.

The increased area proportion of the 'Very High' category (12.8%, tab. 12 to 13.6%, tab. 13) also showed a marginal increase in the prediction of known sites, from 44.2% (tab. 12) to 46.8% (tab. 13). This change altered the mostly the 'High' predictive category, decreasing both the area size (39%, tab. 12 to 38.3%, tab. 13) and number of sites (27.4%, tab. 12 to 25.6%, tab. 13). Overall, the addition of the buffered town areas could be seen as an improvement in the accuracy of the 'Very High' category, however it cannot be known if this reflects the location of unknown Roman sites within Hertfordshire from the results in Table 13 alone.

Table 13: The total count of sites (n = 3466) and area within each predictive category in the Roman Hertfordshire predictive model.

Predictive value	Number of known sites	Site %	Area (km^2)	Area %
Very High	1621	**46.8%**	223.869	**13.6%**
High	886	**25.6%**	630.337	**38.3%**
Medium	792	**22.9%**	579.485	**35.3%**
Low	167	**4.8%**	194.555	**11.8%**
Very Low	0	**0.0%**	15.517	**0.9%**
Total:	3466 sites	**100%**	1643.758 km^2	**100%**

5.2. Evaluating the Predictive Model

The process of testing an archaeological predictive model insinuates that there is a predefined measurement of what a 'good' model is, available for comparison. This is not entirely the case as it is difficult to decide the requirements of a good model without its context and purpose in mind (Verhagen & Whitley 2012, 84). Certain 'universal' requirements of a good predictive model were stated by Verhagen (2009), whom includes features that go beyond testing of performance. Verhagen states that a good model provides an "explanatory framework" for the observed site densities, is "transparent" in its model-building, provides the "best possible prediction" with the current data set while is also able to "perform well in future situations" and specifies "the uncertainty of predictions" (Verhagen 2009a, 63). Only two of these criteria relate to the performance of the model, while the majority of requirements address the way the model is published. This suggests that what is meant by a 'good model' goes beyond its abilities, but still remains an important aspect in its evaluation. The performance of a predictive model is able to be tested through various means, such as expert judgement (Verhagen 2009b, 74), independent data sets (Verhagen 2009b, 78), split sampling (Verhagen 2009a, 65) or statistical methods (Verhagen 2009b, 78).

5.2.1. Split sampling

According to Verhagen and Whitley (2012), a good predictive model displays "repeated consistency between model output and measured response" (Verhagen & Whitley 2012, 84). Split sampling is a method in which a random portion of the known dataset is withheld from model-building in order to validate its predictions (Verhagen 2009a, 65). It is not a perfect method for validation purposes as it does not use truly independent data, but is able to show whether a model can remain consistent with new data (Verhagen 2009b, 92). Only 80% of the dataset (*n* = 3466) was used in the model-building phase of the Roman Hertfordshire predictive model. The remaining 20% of dataset (*n* = 892) can subsequently be used to see the consistency in the model's predictions.

Appendix 22 shows the final predictive model with the entire sample of known archaeological sites (*n* = 4,358). Remarkably, the test data appeared to have a better accuracy rate than the model data it was partially built from (tab. 14). Fewer known sites appear within the 'Very High' category of the model data (46.8%, tab. 14) than in the test data (54.9%, tab. 14), reaching the threshold of predicting over half of the sites within the highest category. This result suggests the model is able, and somewhat accurate, in predicting archaeologically-known Roman sites which were not used to create the model. However, a further way to test the model would be to collect entirely new independent data from surveying or trial trench testing. This form of testing also disregards the precision of the model. The ability to predict 40% of sites within 20% of a given area is more meaningful than predicting 50% of sites within 40% of the area. For this reason, an alternative statistical method is more effective to assess the model's performance.

Table 14: Testing the accuracy of the Roman Hertfordshire predictive model using both the model sample (n = 3466) and the test sample (n = 892).

	Model Sample (80%)		Test Sample (20%)	
Predictive value	**Number of known sites**	**%**	**Number of known sites**	**%**
Very High	1621	**46.8%**	490	**54.9%**
High	886	**25.6%**	173	**19.4%**
Medium	792	**22.9%**	200	**22.4%**
Low	167	**4.8%**	29	**3.3%**
Very Low	0	**0.0%**	0	**0.0%**
Total:	3466 sites	100%	892 sites	100%

5.2.2. Kvamme's Gain

There are various statistical methods that attempt to assess a predictive model's performance, including the K_j parameter (Verhagen 2009b, 76), gross error or wasteful error tests (Verhagen 2009b, 113) or Kvamme's Gain equation (Kvamme 1988, 329). These statistical tests aim to determine the 'performance' of a predictive model by combining its accuracy and precision – two criteria of a 'good' model (Verhagen 2009b, 112). In combination with each other, accuracy ensures there is a high rate of correct prediction, while precision ensures the model's ability to limit the areas of high archaeological probability "as narrowly as possible" (Verhagen 2009a, 64). Model performance criteria is not defined (Verhagen 2009a, 64), so it is assumed that for use in heritage management a model needs both qualities.

Kvamme's Gain is the "most widely used method" to test model performance (Verhagen 2009b, 76) by combining the criteria of accuracy and precision into one comparable 'gain' score (Verhagen 2009a, 64). The equation divides the percentage of the area a predictive zone covers by the percentage of the site count within the zone (Kvamme 1988, 329). The answer is then subtracted from 1 to provide the 'gain' of the predictive zone. A 'high gain' can result from a predictive map whose high predictive areas are small, but includes a large proportion of the site count within them. Low probability zones therefore should have a 'low gain' (Verhagen 2009b, 76). The ideal distribution of gain scores would therefore be the highest (closest value to 1) for high predictive categories and the lowest (below -1) for the lowest predictive categories.

$$\text{Kvamme's Gain} = 1 - \frac{P_a}{P_s}$$

P_a = the area proportion of the zone of interest
P_s = the proportion of sites found in the zone of interest

To validate the final product of the Roman Hertfordshire predictive model, the unused test sample of data was used to produce Kvamme's Gain scores for all of the predictive categories. This provides further validation of the model because it was not built with this sample of data in mind. In addition to this, a Kvamme's Gain score was calculated with the model-building sample as well as with the full data sample. This provides a comparison of performance depending on the data applied to the predictive model.

Table 16 shows both the proportion figures on the left-hand side, with the area proportion in white and the site count proportion in light grey. The test sample produces the highest gain figure for the 'Very High' predictive category (0.75, tab 15), and the lowest levels of gain for the rest of the categories: -0.98, -0.57 and -2.64, respectively. Therefore, the model is validated to work with a small sample of unused data. The 'Very Low' category was unable to produce gain figures with any data sample as the proportion of sites counted remained at zero. The gain remained high for the full data sample (0.72, tab 15), further demonstrating the accuracy and precision of the final model.

Table 15: Comparison of the Kvamme's Gain scores for the final Roman Hertfordshire predictive model using the model sample, test sample and entire sample. Top left of column: area proportion. Bottom left of column: proportion of sites. Right column: Kvamme's Gain score.

	Kvamme's Gain					
Predictive value	Model sample (n= 3466)		Test sample (n= 892)		100% sample (n= 4358)	
Very High	0.14 0.47	0.71	0.14 0.55	0.75	0.14 0.48	0.72
High	0.38 0.26	-0.50	0.38 0.19	-0.98	0.38 0.24	-0.58
Medium	0.35 0.23	-0.54	0.35 0.22	-0.57	0.35 0.23	-0.55
Low	0.12 0.05	-1.46	0.12 0.03	-2.64	0.12 0.04	-1.63
Very Low	0.01 0.00	-	0.01 0.00	-	0.01 0.00	-

For the purpose of transparency, Kvamme's Gain was also calculated from the full data sample for each of the three model builds (tab. 16). The comparison of each model by their gain scores shows the progression of the model's performance throughout its creation. The table clearly demonstrates that Model 2 and Model 3 resulted in the highest gains for the 'Very High' value, and the lowest gains for the 'Very Low' value. This supports the claim that these two models perform the best in terms of accuracy and precision.

Comparatively, Model 3 showed a decrease in gain score of the 'High' category at -0.58 from the score of -0.46 from Model 2. This may have occurred due to the slight decrease in sites observed within the 'High' area without the area size increase equally. The decrease in score could therefore indicate the 'High' category in Model 3 is less accurate than in Model 2, calling into question whether the alterations made within Model 3 can be justified. However, it should be kept in mind that these scores may not be representative of how new, independently-collected data would impact the performance of Model 3.

Table 16: Comparison of the Kvamme's Gain scores for Model 1, Model 2 and Model 3 (the Roman Hertfordshire predictive model) using the entire sample (n = 4358). Top left of column: area proportion. Bottom left of column: proportion of sites. Right column: Kvamme's Gain score.

	Kvamme's Gain					
Predictive value	Model 1		Model 2		Model 3	
Very High	0.14	0.70	0.13	0.72	0.14	0.72
	0.47		0.45		0.48	
High	0.39	-0.49	0.39	-0.46	0.38	-0.58
	0.26		0.27		0.24	
Medium	0.28	-0.62	0.35	-0.52	0.35	-0.55
	0.18		0.23		0.23	
Low	0.18	-0.83	0.12	-1.63	0.12	-1.63
	0.10		0.04		0.04	
Very Low	0.01	-	0.01	-	0.01	-
	0.00		0.00		0.00	

5.3. Applications of the Roman Hertfordshire Predictive Model

The Roman Hertfordshire model, through its presentation with the full sample of known Roman sites (appendix 22) or with modern roads (appendix 23) is meant for use by developers and spatial planners who are attempting to understand the archaeological situation in an area. The accompanying guide (appendix 24) aims to advise the user on the method of intervention that may be suitable for developments within different predictive value zones. Specific information is given to guide the planning process, taking into account the size, depth and archaeological value of the area. A guide on the intended applications of an archaeological predictive model is needed in order to avoid its misuse and misunderstanding. While a predictive model may look like a map, it is important that the model is not viewed, and used, as if it displays infallible knowledge of archaeological potential. Therefore, an understanding of the methods used to define the predicted archaeological risk zones is imperative to its usage in archaeological heritage management.

The methodological approaches which were used to create the model should be documented, as there are many different general approaches (Kamermans *et al.* 2004, 5). This has been done for the Roman Hertfordshire model in the previous sections. The aim of a predictive model can also impact the intended uses of the model, and can be either correlative or explanatory in nature. Correlative aims usually lead to models which only predict the presence of archaeological material in the present, while explanatory aims include understanding, or explaining, human behaviour observed in the past (Kamermans *et al.* 2004, 6; Van Leusen 2002, 102). The predominant aim of the Roman Hertfordshire predictive model was to create a model to be used for archaeological heritage management, and therefore a correlative aim was sufficient for this purpose. However, the model can also be used academically for explanations of site patterns which it can provide through proximity analysis.

5.3.1. Proximity-based analysis

Archaeological predictive models may not only serve the purpose of predicting the level of archaeological risks in the modern landscape, but also may be used to explain patterns in human behaviour within the landscape (Kamermans *et al.* 2009, 10). This can be done by analysing different environmental and social factors in more detail to understand trends which occurred in the data. Analysis can be conducted on the basis of site function or the historical period being investigated (Danese *et al.* 2014, 43). In professional predictive models, many time periods are included so analysis of each period is important to understand period-specific patterns. Analysis by site type is also very important, as certain site functions require different needs from the landscape. According to Verhagen (2009), all good models should provide this explanatory framework for the observed site densities, regardless of their academic uses (Verhagen 2009a, 63) as it can usually explain the factors which influenced the predictions. However, this kind of academic assessment can run the risk of becoming too inferential as analysis heavily relies on the known data being representative.

As only one historical time period was included in the Roman Hertfordshire predictive model, this type of in-depth analysis of the period has already been achieved. However, sites can be analysed by their specific function, such as settlements, economic, ritual or military sites, in order to analyse the potential influence of landscape factors on their location. The landscape factors which were analysed for influence include water access proximity, road proximity, elevation and the proximity to the Roman city of Verulamium. These factors were chosen to represent how both social and environmental could affect different site types.

Analysis of each factor was conducted through the use of the 'Point Sampling Tool', installed as a plug-in within QGIS. This tool allowed the quick collection of proximity values of point data to any given raster layer. A rasterized layer of the water sources and roads were reused from the model building. To calculate distance from the city of Verulamium, a 2km buffer area was created around the central point of the city and was rasterised. The elevation was already provided in a raster format within the EU DEM. All four raster layers were used to create four proximity rasters without a specified limit of proximity that would then be used in the Point Sampling Tool. The point data used in this analysis included the model-building data sample (n = 3466) as they were previously grouped by site type. However, since only settlement, economic, military and ritual site types were included in the analysis, there was a total of 2708 sites used.

The tool selected the proximity rasters of the water sources, roads and city of Verulamium and the data sample of each site type group individually, creating a list of proximity values from the subject. The elevation raster was also used in this tool, creating a list of elevation values at each point. The data for each input raster was copied into a Microsoft Excel spreadsheet, where values were organised by their site type (fig. 59). Minimum and maximum values were determined and the appropriate class limits were created in order to produce frequency charts and graphs. Data for each of the four factors were displayed on the same graph, making for a visual comparison between site types.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
			settlements			economic			military			ritual	
max:	class limits		ogc_fid	distance		ogc_fid	frequency_		ogc_fid	frequency_		ogc_fid	frequency_
12184.00586	500		617	12184.00586		304	282.84271		53	2236.06787		304	1923.53845
min:	1000		1456	11704.69922		303	2529.82227		52	141.42136		303	282.84271
0	1500		1317	11704.69922		302	141.42136		51	4652.95605		302	854.40033
	2000		383	11704.69922		301	2758.62305		50	6596.96875		301	5813.77686
	2500		616	10643.77734		300	1910.49731		49	412.31055		300	6152.23535
	3000		1725	10643.77734		299	5597.3208		16	11704.69922		299	300
	3500		1692	10643.77734		298	282.84271		15	4661.54492		298	200
	4000		1645	10643.77734		297	761.57733		14	223.6068		297	2529.82227
	4500		358	10486.18164		296	721.11023		13	4242.64063		296	141.42136
	5000		705	10295.62988		295	7467.26172		12	200		295	316.22778
	5500		67	10295.62988		294	2900		11	360.55511		294	300
	6000		317	10285.91211		293	1910.49731		10	200		293	2236.06787
	6500		68	10155.29395		292	100		9	200		292	5457.10547
	7000		618	9391.48535		291	282.84271		8	1264.91113		291	1264.91113
	7500		1419	9391.48535		290	4738.14307		7	500		290	500
	8000		1067	9391.48535		289	761.57733		6	4560.70166		289	1700
			1614	9391.48535		320	2236.06787		5	100		318	282.84271
			511	9386.16016		319	2236.06787		4	0		317	0
			588	9347.19238		318	5597.3208		3	1581.13879		316	1500
			956	8402.38086		317	1910.49731		2	5434.15088		315	632.45557
			1492	8402.38086		316	1910.49731		1	500		314	282.84271
			1579	8402.38086		315	3613.8623		32	4177.31982		313	700
			19	8402.38086		314	2529.82227		31	412.31055		312	412.31055
			1043	8049.84424		313	141.42136		30	2915.47607		311	670.82037
			368	8000		312	5597.3208		29	3400		310	500

Figure 59: View in Microsoft Excel of the listed distances (m) from the Roman roads to sites of different types (settlements, economic, military and ritual).

Frequency analysis showed that sites of all types were influenced by the location of the Roman roads (fig. 60). The graph suggests that military sites may have been the least influenced by the location of roads, perhaps due to a higher priority being given to more strategic locations away from the populated roads. However, while the role of Roman roads in military campaigning within Roman

Britain is not well understood, it is likely that the Roman armies in the field did not directly depend on roads for transport (Menard 2011, 41). The suggested lack of road-influence could also have been due to the lack of identifiable military sites (*n* = 53) in proportion to other site types. Settlement sites are represented the most in the known data, but also appeared to have the strongest connection to road proximity in the proximity graph.

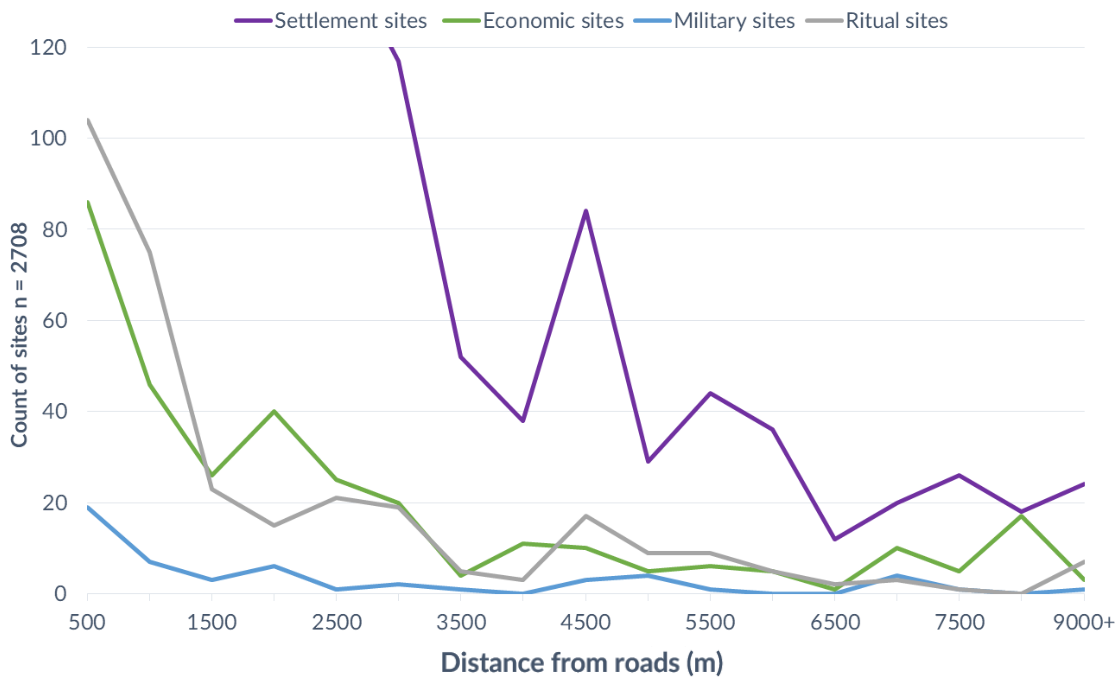


Figure 60: Frequency graph showing the count of sites (according to site type) at various distances (m) from the Roman roads (n = 2708).

Within 500m of the roads, there were 628 settlement-related sites, which decreased to 450 sites between 500m and 1000m. This number decreased the longer the distance of the roads became, with a small increase at 3500-4000m. Economic and ritual sites appeared to decrease in site count at a similar rate and distance from roads, suggesting that both economic and ritual sites required transport routes more so than military sites within the Roman period. In regards to the ritual sites, it is known that Roman law required most burials to be located outside the city walls and therefore often became concentrated "on the approach roads" outside the city (Historic England 2018, 8). Economic sites would have also needed direct routes to trading centers within Roman towns and cities.

Within the data, distance from water appeared to be even more of a stronger influence to all site types than the distance from roads (fig. 61). Site distances from water sources beyond 6000m appeared to be non-occurring in the data. Similarly to road proximity, military sites appeared to be the least influenced by water access among the other site types. However, this could also support the idea the sample size of the military sites is too small to make conclusive judgements. Economic and ritual sites also remained similar in their proximity to water sources, which could be more related to their often close proximity to towns which would have considered both road and water access. Settlement sites were strongly correlated to close proximity to water, with over 50% (1236 sites) of settlement-related sites being located within 500m from any water source.

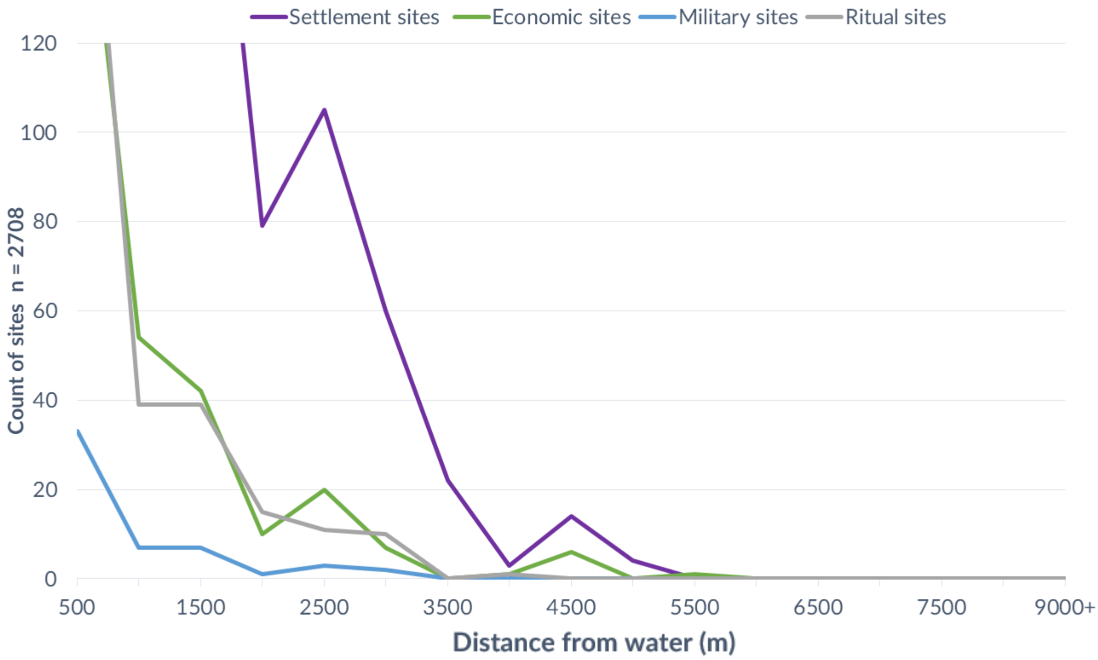


Figure 61: Frequency graph showing the count of sites (according to site type) at various distances (m) from water sources (n = 2708).

Most sites were either in close proximity to Verulamium, or were above 8500m away (fig. 62). The high proportion of sites of each type within 1000m were likely the only sites influenced by the location of Verulamium. The economic and ritual sites were likely connected to the city in some manner, whether it was for trade, worship or for burial sites outside of the city walls. The other cluster of sites further away from Verulamium were likely connected to other Roman towns within Hertfordshire, such as Baldock, Braughing, Ware and Welwyn (Tereszczuk 2004, 10). With the creation of road systems, close proximity to certain centers may not have been as much of a priority as roads would provide travel links to many centers. The catchment area of influence therefore grew with the creation of zones, making a town's location near these roads of high importance.

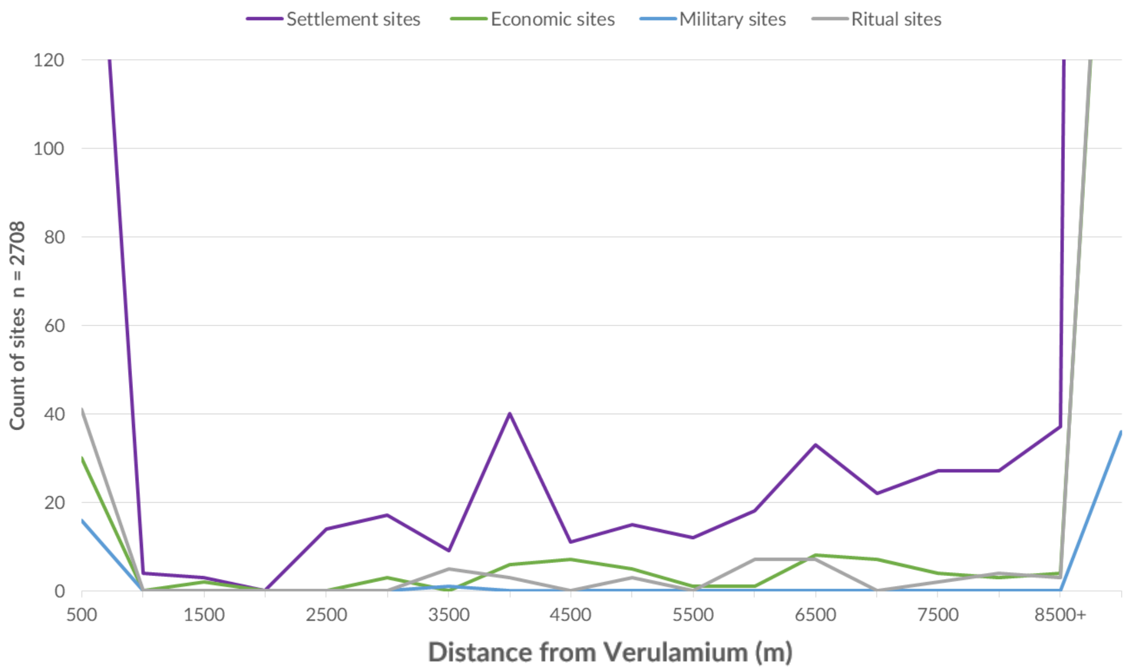


Figure 62: Frequency graph showing the count of sites (according to site type) at various distances (m) Verulamium (n = 2708).

The elevation of all of the site types were most frequently between 75-100 meters above sea level (fig. 63). This is clearly the case for known settlement sites, where over half of sites were between this range. Economic and ritual sites had a shallow curve within elevation levels, however also peaking in frequency at the 75-100 meter mark. This distribution was likely affected by the modern elevation data used, as Hertfordshire appears relatively flat except for slight elevation within the Chiltern Hills in the North of the county. The elevation of the terrain was likely not a highly contributing factor in site location in the Roman period within this area of England, as it remains to be relatively low in elevation.

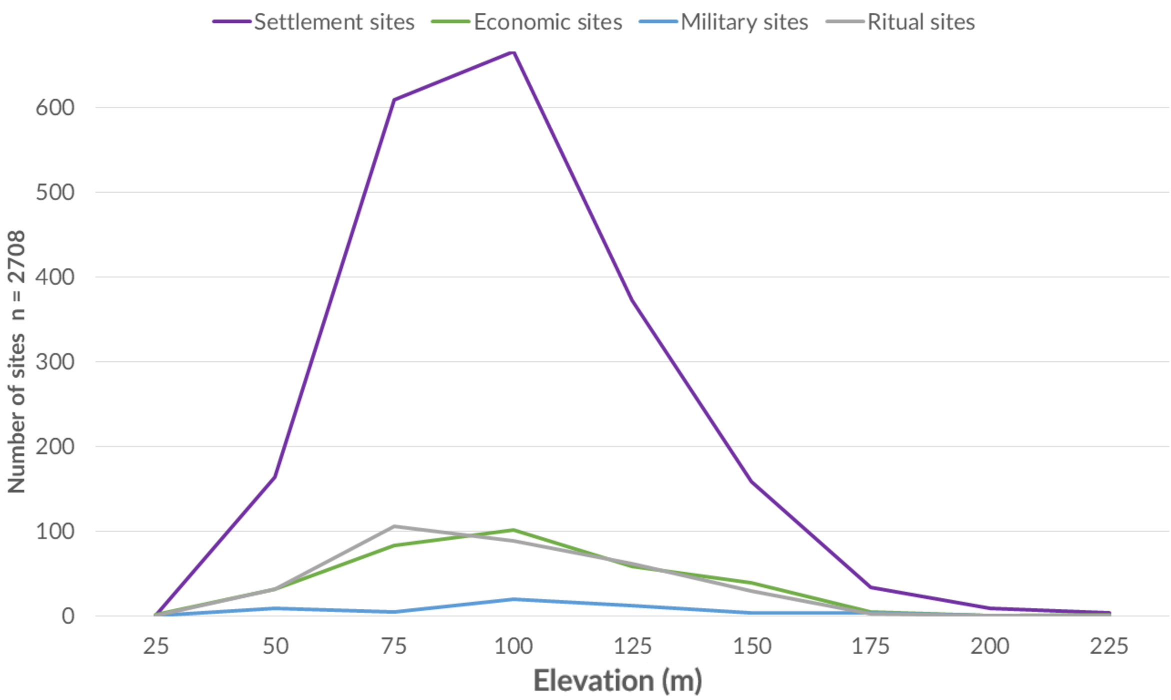


Figure 63: Frequency graph showing the count of sites (according to site type) at various elevations (m) (n = 2708).

6. Discussion

Archaeological predictive modelling is a tool of convenience. Its implementation within heritage management can lead to quicker decision-making in situations of spatial planning, therefore lessening the risk of disturbance of archaeology. However, a strong benefit of predictive modelling is the potential it has for saving money and time on excavations that may be deemed excessive. Inductive predictive maps have been produced and used quickly and simply within the Netherlands for more than 15 years (Verbruggen 2009, 28). This has been possible due to the availability of an "extensive dataset" of archaeological find spots (Archis), detailed soil maps and the use of the GIS application. However, it has been stated that the Dutch national predictive maps are "heavily distorted by an overrepresentation of sites on or close to the surface" (Verbruggen 2009, 28).

It should not be denied or dismissed that predictive modelling is prone to serious issues if assessments are not made of the methods and data used in the production of the model. This is because, by the very nature of prediction, our assumptions of the unknown can only ever come from the known.

Before implementation of predictive modelling could successfully be integrated into the English AHM system, the wider issues of funding and standardisation of archaeological predictive models should also be discussed.

6.1. Guidance for the Roman Hertfordshire Model

The Roman Hertfordshire predictive model aims to predict site locations in order to protect unknown archaeology. The main way in which the model can achieve this is to also 'predict' how much intervention is needed before construction in order to prevent unnecessary damage, providing transparent decision-making (Lauwerier *et al.* 2018). Archaeology must be dealt with in the most efficient way by both the developer, archaeological contractor and the authorities in order to achieve this goal (Kamermans *et al.* 2009, 10). For these reasons, a guide (appendix 24) was made to accompany the Roman Hertfordshire predictive model to suggest appropriate methods of intervention on the basis of the size of development and the predictive value of the area. However, a discussion should be had on the thought process behind this guidance, and the potential alterations which could improve them in the future.

Ideally, predictive models would aid in the decision-making process within the AHM system by providing more direct guidance than is already suggested by the current documentation (Kamermans *et al.* 2009, 10), such as can be found within the 'Historic Environment Good Practice Advice in Planning' from 'English Heritage (Historic England, since 2015). This can be achieved through a series of 'rules' for action depending on the predictive value of the area, size of the proposed development and the depth of soil disturbance required. Continued improving of our knowledge on the extent and depth of soil disturbance can lead to a more responsible and efficient approach to these issues (Lauwerier *et al.* 2018) in addition to the refinement of a predictive model over the course of time. Therefore both the current state of the model and its accompanying guide are subject to updates and alterations over time.

The guidelines for methods of intervention would be applied in situations where the development project is likely to disturb the top soil layer, which is typically 30 cm in depth. This can be from digging, laying foundations, or the placement of a load-bearing layer which can alter the soil layer which contains archaeology (fig. 64). Development projects which would not disturb 30cm below the ground may not need any form of intervention as it would not pose a high risk of archaeological disturbance.

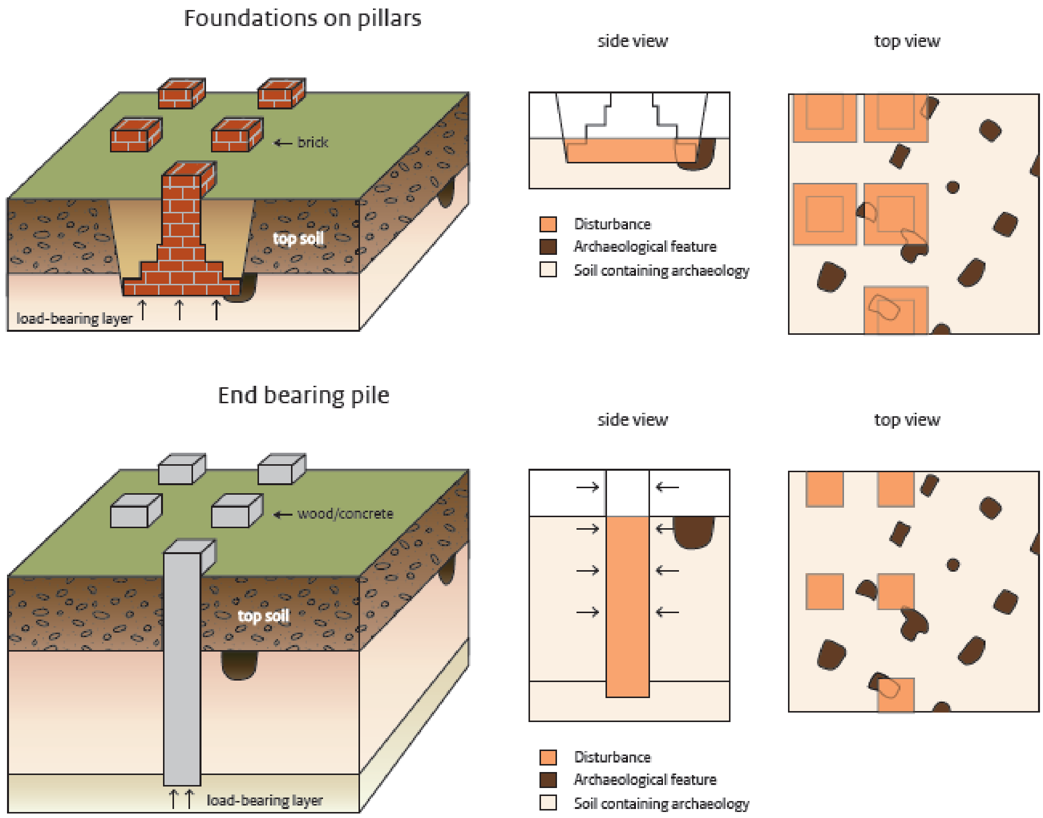


Figure 64: Diagram of the alteration of the underlying archaeological layer due to the laying of foundations or load-bearing layers (after Bouwmeester et al. 2017, 150).

It is assumed in the guide (appendix 24) that developments which are intended to be small, in accordance to their predictive value areas, would not pose a high level of risk to archaeology. The advice placed great importance on the assignment of an area's given predictive value. A development of over 50,000 m^2 without any more of an intervention than a desk assessment is dangerous if the values are completely incorrect. It is for this reason that even the 'Low' valued areas should perhaps require both a desk assessment and core sampling procedures before planning permission would be granted.

Certain alterations could be made to the rules to differentiate between areas with urban or rural modern land use which could account for potential disturbances that have already occurred. However, it is better to err on the side of caution when assuming which soils have or have not already been disturbed as there is no suitable method to estimate the extent of disturbance (Lauwerier *et al.* 2018).

Within the higher valued areas, perhaps more lenient methods of intervention could be suggested for developments that are below the size restrictions as it could lead developers to use these sizes as a standard to entirely avoid any extra expenses at the hands of archaeologists. However, it should be stated that if any remains of archaeology are found at any point during the advised method of intervention, further action must be taken. The levels of preliminary intervention may only help developers gauge an idea of how much expenses should be expected for archaeological excavations within a given area.

In situations where archaeological remains are found, an evaluation of the preservation of the discovered archaeology should be undertaken by specialists. If the archaeology is deemed to be in reasonably good form, the developer must decide if preservation *in situ* can be accommodated in the construction plans or if the development can take place elsewhere. If the remains are unable to remain *in situ*, or the preservation is not deemed to be good, excavation of the material should take place within the area of development.

6.2. Funding for Predictive Modelling

The answer of funding the creation of either regional or national archaeological predictive models is paramount to their potential adoption within the English AHM system. The cost-saving benefits of predictive models within heritage management is a large benefit of the method (Verhagen & Whitley 2012, 50; Verhagen *et al.* 2014, 379). Within the Netherlands, successes were first seen in the form of small-scale, broad inductive models (Verbruggen 2009, 27). These would have been dramatically cheaper and easier to create than the large-scale detailed deductive models that are now created within regions of the Netherlands (Verbruggen 2009, 27). This period of 'adjustment' allows for the trial of archaeological predictive modelling without significant costly attachments to the method. However, if large-scale models are created there would be a short-term rise in costs, with some amount of years of use needed before the costs would have paid-off. In 1999, the state of Minnesota in the U.S. spent $4.5 million on the creation of the Minnesota archaeological predictive model, but it is now said to be saving the U.S. state roughly $3 million per year as compared to the previous system of heritage management (Wilcox 2014, 342). Within England, an alternative candidate for the funding of predictive models, besides the state, could be the development company themselves, of which the models could be reused after the specific instance. This is possible as developers are already expected to fund an archaeological investigation, decided by the local authorities (Wilcox 2012, 353).

However, the application of predictive modelling varies drastically depending on the scale, detail and archaeological situation. It is possible that the application of such a method would not be the same in any two places. It is possible that predictive models would only need to be reinstated in certain parts of England, such as is the case within the Canadian wilderness regions. Areas within Canada that are known to lack archaeological data are modelled for their predictive capabilities, which is especially used in the large forested areas (Wilcox 2014, 343). Within England, this selective method of modelling would likely impact areas outside of Hertfordshire, between urban areas within large rural landscapes. However, this could also be applied to areas where archaeology is even more at risk, such as within areas of off-shore mineral extraction off the coast of the North Sea and the English channel (Wilcox 2014, 343). Due to modern sea-level rising, much of the coastal areas inhabited once by man are under the water off of these coasts. Mineral extraction often threatens these areas, and archaeological remains can be encountered within the various strata of a minerals extraction site (Historic England 2020, 12). The level of preservation can also be extraordinarily high, such as was the unsuspected case at Must Farm Quarry, Cambridgeshire, where waterlogging and charring preserved a Late

Bronze Age settlement of round stilted wooden structures (Historic England 2020, 12). It is possible that these chance discoveries do not have to be accidental, with the guidance of predictive modelling in unrecorded, at-risk areas (Wilcox 2012, 353). With smaller predictive models, there is the chance to fund it privately. Within the Netherlands and Canada, smaller models are produced by private companies and then funded by the developer (Wilcox 2012, 356).

One issue is how much to fund predictive models, but another issue regards who should fund them and which consequences this may have on the resulting models. Kamermans *et al*. (2009) remarked that it is “undesirable” that archaeological predictive models should be produced by archaeological companies (Kamermans *et al.* 2009, 11). At a surface level, it would make sense that archaeological companies are in charge of using expert judgement and archaeologically-related theories to produce archaeological predictions. However, it is possible that with the underlying knowledge that a larger area with high archaeological predictions generates more excavation work the intention to increase profits and control can be realised (Meffert 2009, 33). Therefore, the responsibility must be handed to people who are least likely to gain from motivated modelling. In the Netherlands, the creation of predictive models are the responsibility of ‘independently operating’ government institutions (Kamermans *et al.* 2009, 11) such as the Cultural Heritage Agency of the Netherlands (RCE), formerly known as the RACM (The National Service for Archaeology, Cultural Landscape and Built Heritage) (Kamermans *et al.* 2009, 12). The RCE can profit from larger areas of high archaeological value, however since it is operated by the state there are measures in place to keep motivations as neutral as possible. As well as this, local authorities should look further at aspects of site type rarities, nature and quality of remains and landscape genetics in addition to the predictive model values.

Ideally, a specialised government department of cultural heritage within the UK would be the most obvious candidate for the creation of national or regional predictive models. However, the department which handles heritage is the

Department of Culture, Media and Sport (DCMS) since 1997, formerly the Department of National Heritage (Benetti & Brogiolo 2018, 179). The DCMS allocated funding for archaeological heritage management to non-governmental institutions like Historic England (English Heritage) or The Royal Commission for Historic Monuments. Historic England could be a contender for predictive modelling within England, but is not independently operating from the free-market. Perhaps a remedy would be for the creation of predictive models by Historic England, and a checking phase would be undertaken by a senior member of the DCMS before approval of the predictive model is granted.

6.3. Reproducibility and 'Open Science'

The data which was collected to inform and produce the Roman Hertfordshire predictive model was entirely possible due to the open accessibility of archaeological and environmental data from internet sources. A majority of the digital data was repurposed to use within the modelling process, such as the reclassification of environmental layers and the data cleansing of the archaeological site data from the Archaeological Data Service (ADS). This was only possible because of the publication of this digital data by various sources, most significantly from the digital archaeological repository collated by the ADS. Research on the re-use of ADS data was conducted by Huggett (2018), and found that while it was difficult to evaluate the level of re-use (Huggett 2018, 94) it was commonly repurposed in a way in which the data was not originally collected for (Huggett 2018, 96). The internet allows for a wider access of such recorded data and thereby creates new opportunities for a variety of interpretations.

Open licensing of data also promotes re-use and repurposing, which was the case all of the environmental sources used in this project. For example, the 'BGS Geology' model by the British Geological Survey was provided in four forms, varying by their price and licencing. The lowest resolution form of the model ('BGS Geology 625k') was used in the project as it is "free for commercial, research and public use under the Open Government Licence" (www.bgs.ac.uk/products/digitalmaps) with an acknowledgement of the creators of the model.

The citing of re-used data sources are also a matter of discussion in the 'Open Science' debate, with no agreed-upon standard on how to properly and effectively acknowledge the archaeological data used. Huggett (2018) stated a reason for the "lack of means to evaluate levels of re-use" of ADS data was due to the limited or informal manner that digital data is cited (Huggett 2018, 95). An ethical consequence of this is a "perceived lack of credit" that may accompany the sharing of archaeological data by researchers (Marwick & Pilaar Birch 2018, 1). The preparation, costs and lack of standards involved in digital data sharing would likely dissuade individual researchers from considering the option (Marwick & Pilaar Birch 2018, 4). An incentive for an individual researcher to publish their data online can be partly intangible, with the aim that it can be used again to increase their work's productivity. However, ensuring appropriate credit is given to the researcher provides a more tangible benefit to the prospect of sharing their data (Marwick & Pilaar Birch 2018, 3).

A standard for citing data could be achieved through the promotion of DOI usage (Digital Object Identifiers) for digital data citing. However, the use of DOIs are not always suitable to cite many individual records, as is the case when exporting a database. For example, the ADS repository lists a DOI in association with each ADS Archive and report and therefore provides a more stable source of credit than a regular URL (www.archaeologydataservice.ac.uk/about). Use of their records requires acknowledgement of its DOI, however the repository was used within this research to export an entire dataset of records. The selection was made by filtering the location and time period, overall providing the use of 4358 records that were provided by a range of researchers. Neither the ADS' Q&A page nor the Leiden University guidelines suggested a manner in which to reference thousands of records besides adding the URL of the 'ArcSearch' webpage to the 'Internet Sources' section at the end of this research. This is perhaps not the most suitable way give appropriate credit as the metadata and individual providers of each record are unable to be cited.

The open accessibility of the data layers gathered from various established sources (Ordnance Survey, British Geological Survey, UK Government, University of York and Harvard University) also ensures that the Roman Hertfordshire predictive model could be reproduced in the future, using either the same or alternate methods. This is an important aspect of digitally-based archaeological research as open accessibility, and thus reproducibility, ensures the opportunity for improvement.

6.4. Standards for Predictive Modelling

Each predictive model is as unique as the landscape it was modelled from, which is likely a contributing reason for little generalised standards for the creation or output of predictive models. It can be assumed that almost all kinds of archaeological predictive models will include defined areas of an assigned value that was either inductively or deductively determined, with a legend explaining what each colour-coding indicates on the final map. Beyond this assumption, the presentation and publication of predictive models vary greatly. This is an alarming issue for predictive modelling, as certain digital standards and metadata are needed for continued use of the model. It should therefore be discussed the possibility of forming standards for predictive models, if their implementation would enter the English AHM system.

There is also the option of forming 'best practices' for predictive modelling. This is often proposed in opposition to standards, which is a form of template or a guideline that creates a basis of comparison. Best practices can be defined as a method or technique that is the most accepted, or is prescribed as being the most correct or effective. While both seek to control the quality of the modelling process, standardisation is perhaps most suitable for archaeological predictive modelling due to the varying nature of the task. The location of the research area, temporal or geographical scale, data quality, budget and experience of the researcher can all have large impacts on the modelling process and outcome. This would make it unrealistic to assume one modelling method would be the most effective for all predictive modelling projects. With that being said, a general level of uniformity in the presentation or evaluation of archaeological predictive models would be beneficial to the method. For example, it is widely agreed upon that models should be tested (Nakoinz 2018, 105; Wilcox 2014, 344; Verhagen 2009a, 63; Verhagen & Whitley 2012, 83), and therefore there could be a development of standards that include the publishing of test results alongside the model.

6.4.1. General standards

The standards that should be applied to all predictive models within a given country should first follow a general standard which includes basic information that must always be included in a model's approval and publication. Within British Columbia, Canada, these basic standards (Wilcox 2014, 345) exist as a suggestion for their application in other countries who also use predictive modelling in their heritage management.

Firstly, technical standards must be maintained which include a national map projection for each model and a suitable, universal file format. Both of these elements are crucial to avoid a situation where two independently-made models are produced in alternate projections and cannot be merged, such as can be seen during the creation of the national Dutch Indicative Map of Archaeological Values (IKAW) model (Meffert 2009, 33).

Secondly, quotes of scores are expected to be provided for the Kvamme's Gain expression, and archaeological site density within each category. This could then ensure an extra level of quality assurance upon publication, in addition to providing context on risk assessment (Wilcox 2014, 345).

Metadata about the author and their contact details ensures the accountability of the model's predictions as well as an opportunity to open up ways of contact upon further reuse of the model. Contextual metadata should be included alongside the model, such as descriptions of the geology and topology of the area, and an assessment of the archaeological and environmental data that was used. Transparent descriptions of the methods and theory behind the predictive values is also needed to increase understanding of the model and repeatability (Wilcox 2014, 346). Perhaps a research mask of the area could also be included over a display of the model, to visually account for the non-researched areas.

6.4.2. Standards for predictive modelling for AHM

General standards included basic information which must be given alongside the model. The standards for the act of predictive modelling for Archaeological Heritage Management (AHM) should detail different expectations within the process of modelling. The Dutch Archaeological Quality Standards (KNA) (Willems & Brandt 2004, 17) details the standard process of creating regional predictive models for desk-based assessments within heritage management. It provides some suggestions for a standard system of predictive modelling which could be integrated into the current English AHM system.

The models are created by the collaborated effort of junior, mid-level, and senior archaeologists (Willems & Brandt 2004, 28). Firstly, information should be requested for the background and creation of the model. Within the KNA, this information includes the national Dutch IKAW, geographical maps, soil maps, geomorphological maps, contour maps and remote sensing maps (Willems & Brandt 2004, 41). Modern maps on current land-use, historical maps or groundwater data should also be found to aid in the modelling process. No specific method of modelling is recommended within the standards, which is likely because the most suitable method would be dependent on the size, terrain and available data. Further information is requested from the developers, detailing the level of disturbance within the short and long-term. Short-term information is gauged by questions about the nature and size of the proposed development, the depth and method of soil removal required, the amount and location of displaced soil. Long-term information is gathered from questioning about the eventual creation of watercourses and pavements, as well as future plans of the development (Willems & Brandt 2004, 37). The model is then created, usually by a Junior Archaeologist who assigns areas and the expected archaeological values to the research area (Willems & Brandt 2004, 28). This version is externally checked by a Senior Archaeologist who signs off the map to the developer seeking planning-permission (Willems & Brandt 2004, 35).

This process of modelling could be improved by including a standard which recommends certain actions to be taken within differently valued areas. This would be to ensure the developer and local authorities use the predictive value of the model in the way it was intended. Through the integration of these general standards and modelling process standards, a stream-lined production of well-contextualised predictive models could be created for use in English Archaeological Heritage Management.

7. Conclusion

To conclude, the Roman Hertfordshire predictive model acts as a case study for the collation and use of English-based open data in predicting archaeological site patterns. This case study aims to assess the potential of archaeological predictive modelling within the Archaeological Heritage Management (AHM) system in England, whilst also gaining an insight into the archaeological situation of Hertfordshire in the Roman age. These three research questions will be answered, in reference to the analysis and evaluations made throughout this thesis.

1. Does England have the open-access digital infrastructure to facilitate the creation of well-informed archaeological predictive models?

The data collection process for Roman Hertfordshire suggested that there is considerable amounts of open-access data for the area of England, provided both by private companies in the UK and the EU as well as by the UK government data services. Both social and environmental factors were able to be integrated into predictions, and multiple layers were acquired for the assessment of the landscape. However, the model was limited due to the low resolution of the soil map layer, in addition to environmental layers that are oriented to represent only the modern Hertfordshire landscape. A reasonably large sample of known Roman archaeological data was accessed from the Archaeological Data Service, but lacks information on areas which have been excavated but no archaeology was found. The inclusion of this information could aid in the creation of predictive models, as well as accounting for observational biases in archaeological data. Overall, this assessment supports the conclusion that England possesses the digital infrastructure needed to facilitate the creation of a reasonably well-informed model, such as the Roman Hertfordshire predictive model. However, higher resolution data sources and alternative archaeological information could have strengthened the predictive abilities of the final model.

2. What knowledge can be gained from the creation and output of the Roman Hertfordshire predictive model?

The creation and subsequent evaluation of the Roman Hertfordshire model allowed for a practical view of the theory and methods used in predictive modelling. The case study only represents the Roman period of archaeology, but the methods can be applied to other time periods in order to produce a better model for AHM. In addition to this, the scale of Hertfordshire proved to be fairly large and so the areas of predicted values were generalised. This wide-scale extent of the research area led to the resolution of certain environmental layers being less of an issue than it would have been if the predictive model was a fraction of the size of Hertfordshire. The evaluation of the model also highlighted the importance of using a testing method which takes into account both the accuracy and precision of the predictive categories. It is very simple to compare two site count proportions and not take into account the proportionate size of the area as a defining factor. Kvamme's Gain provides a simple way to compare and validate these two requirements of a predictive model whilst being understandable to someone who is not experienced with statistical testing.

Archaeologically, the predictive model can only confidently inform about the Roman Hertfordshire site patterns as based on preferred proximity to water, roads and towns and the predicted optimal slope and aspect for solar radiation. Therefore, most archaeological interpretations from this model alone would likely not be conclusive, but it may indicate areas where archaeological knowledge could be gained about Roman-age Hertfordshire. These would include areas near the current-day main rivers and within the vicinity of modern towns such as St. Albans, Baldock, Ware, Welwyn, Braughing and Bishop's Stortford.

3. How can the case study of Roman Hertfordshire assess the potential of archaeological predictive modelling within the Archaeological Heritage Management system in England?

The discussion of the model provides an assessment of the standards that need to be implemented for archaeological predictive models within England if the method was to be integrated into the first stages of the AHM system. Matters of sources of funding the creation of the models continues to pose an opposition to its use in the AHM system, but may ultimately lessen the amount of money wasted within the process of protecting archaeology over the long-term. The parties whom should be responsible for predictive modelling is also unclear, but a strong candidate would be within the organisation of Historic England with an overseer from the Department for Digital, Culture, Media and Sport.

Therefore, after the creation of the Roman Hertfordshire predictive model and the considerations that needed to be made about its interpretation, the technique of archaeological predictive modelling has strong potential benefits for its implementation in the English AHM system. However, further consideration must be taken to create standards for testing and applying these models in order to protect unknown archaeology.

The integral criticisms which predictive models are prone to receive within the England can be temporarily addressed through the creation of these national standards that must be adhered to for creation and publication. In order to address the need for funding, gradual implementation of the method should be enacted within rural and off-coast areas that have small samples of known archaeological sites. The predictive aspect would be the most effective in these areas as current AHM methods which relies partly on prior archaeological knowledge would be less suitable. Any costs associated with the production of the models in areas of little archaeological knowledge would be a worthy endeavour as the model would provide a basis of guidance for off-shore mining and planning permissions. This predictive model could then continue to become more refined as more knowledge is obtained.

8. Abstract

In this thesis, the potential of archaeological predictive modelling within England's Archaeological Heritage Management (AHM) system is assessed through the case study of Roman-age Hertfordshire, in south-east England. The case study involves the creation of an archaeological predictive model from the bottom-up, using only open-access data. An assessment is also made on the quality of the open-access digital infrastructure within England, as well as on the knowledge that can be gained from the creation and product of the model.

A detailed description of the collected data provides information about the environment (elevation, soil, geology, hydrogeology and river system), the social aspects of the landscape (Roman road network and towns) in addition to the modern-day factors which impact planning permissions (land-use, modern roads, protected areas and scheduled monuments). The quality of the environmental data is evaluated for its applicability to the Roman landscape in Hertfordshire. Archaeological site data (*n* = 4358), provided by the Archaeological Data Service (ADS), is categorised into site types (settlements, economic, ritual, military, water sources and miscellaneous). The representability of the archaeological data is evaluated for potential observational biases.

The Roman Hertfordshire predictive model is created using deductive techniques (weighted multi-criteria analysis) and an inductive technique (site density). The final product predicts five areas of archaeological potential within Roman Hertfordshire, ranging from 'Very High' to 'Very Low'. The model is evaluated for its predictive abilities by an unused testing sample of archaeological sites. The accuracy and precision of the model's predictions are tested using Kvamme's Gain equation, producing a high-yielding score of 0.72. The applications of the Roman Hertfordshire predictive model are discussed in the context of its uses within the modern development process. Proximity-based analysis of the different site types is explored in regards to water sources, Roman roads and Verulamium (St. Albans). The elevation of different site types are also analysed.

Internet Sources

https://archaeologydataservice.ac.uk/about/faq.xhtml#UFAQ10, accessed 29 August 2020.

https://www.archaeologists.net/find/clientguide/legislation, accessed 11 June 2020.

https://www.archaeologydataservice.ac.uk, accessed 20 April 2020.

https://www.archaeologydataservice.ac.uk/archsearch, accessed 22 July 2020.

https://www.bgs.ac.uk/discoveringGeology/time/timechart/phanerozoic/cenozoic.html, accessed 9 June 2020.

https://www.bgs.ac.uk/products/digitalmaps/DiGMapGB_625.html accessed 20 April 2020.

https://www.bgs.ac.uk/products/hydrogeology/permeability.html, accessed 10 June 2020.

https://www.bgs.ac.uk/products/onshore/soilPMM.html accessed 20 April 2020.

https://www.cultureelerfgoed.nl, accessed 13 August 2020.

https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/143/signatures, accessed 12 June 2020.

https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/066, accessed 19 May 2020.

https://www.coe.int/en/web/culture-and-heritage/valletta-convention, accessed 19 May 2020.

https://www.darmc.harvard.edu/data-availability, accessed 10 June 2020.

https://www.data.gov.uk/dataset/7ed9cc96-82b3-4e1e-a3d9-be61899045e6/archaeological-areas, accessed 23 April 2020.

https://www.designingbuildings.co.uk/wiki/Planning_Policy_Statement_5_Planning_for_the_Historic_Environment, accessed 12 June 2020.

https://www.geoportal.statistics.gov.uk/datasets/counties-april-2019-en-bfc?geometry=-27.609%2C46.136%2C31.673%2C55.778, accessed 20 April 2020.

https://www.heritagegateway.org.uk/gateway, accessed 12 June 2020.

https://www.historicengland.org.uk/listing/the-list, accessed 12 June 2020.

https://www.historicengland.org.uk/listing/the-list/data-downloads, accessed 23 April 2020.

https://www.historicengland.org.uk/listing/what-is-designation/scheduled-monuments, accessed 11 June 2020.

https://www.land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata, accessed 21 April 2020.

https://www.land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=metadata, accessed 21 April 2020.

https://www.legislation.gov.uk/ukpga/1979/46/contents, accessed 12 June 2020.

https://www.ordnancesurvey.co.uk/opendatadownload/products.html, accessed 22 April 2020.

https://www.plugins.qgis.org/plugins/pointsamplingtool, accessed 22 July 2020.

https://www.pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/understanding-solar-radiation-analysis.htm, accessed 9 June 2020.

https://stratigraphy.org/timescale, accessed 29 August 2020.

Bibliography

Benetti, F., and G. Brogiolo, 2018. Polar Opposites: The Legislative Management of Archaeological Research in Italy and England and the Challenge of Public Participation. *Public Archaeology* 17(4), 176-192.

Bouwmeester, H.M.P., J. Abrahamse, J.E. and A.M. Blom, 2017. Mapping disturbances. Potential disturbance of archaeological remains in built-up areas, in R.C.G.M. Lauwerier, M.C. Eerden, B.J. Groenewoudt, M.A. Lascaris, E. Rensink, B.I. Smit, B.P. Speleers and J. Van Doesburg (eds), Knowledge for Informed Choices. Tools for more effective and efficient selection of valuable archaeology in the Netherlands. *Nederlandse Archeologische* Rapporten 55, Rijksdienst voor het Cultureel Erfgoed, Amersfoort. 143-155.

Brandt, R., B.J. Groenewoudt and K.L. Kvamme, 1992. An experiment in archaeological site location: Modeling in the Netherlands using GIS techniques. *World Archaeology* 24(2), 268-282. https://doi.org/10.1080/00438243.1992.9980207.

Brown, A.G., 1997. *Alluvial Geoarchaeology Floodplain Archaeology and Environmental Change*. Cambridge: Cambridge University Press (Cambridge Manuals in Archaeology).

Carleton, W., J. Conolly and G. Ianonne, 2012. A locally-adaptive model of archaeological potential (LAMAP*). Journal of Archaeological Science* 39(11), 3371-3385. https://doi.org/10.1016/j.jas.2012.05.022.

Chartered Institute for Archaeologists, 2017. Standard and guidance for historic environment desk-based assessment, Reading.

Council of Europe, 1992. Explanatory Report to the European Convention on the Protection of the Archaeological Heritage (Revised), *European Treaty Series - No. 143*, Valletta.

Dacorum Borough Council, 2004. Landscape Character Assessment for Dacorum. *Hertfordshire Landscape Strategy 2004.* Landscape Character Assessment for Dacorum, Hertfordshire.

Danese, M., N. Masini, M. Biscione and R. Lasaponara, 2014. Predictive modeling for preventive Archaeology: overview and case study. *Central European Journal of Geosciences 6*, 42-55. https://doi.org/10.2478/s13533-012-0160-5.

Deeben, J.H.C., W.J.B. Derickx, B.J. Groenewoudt, D.P. Hallewas, J.H.M. Peters and E. Rensink, 2009. *Handleiding voor de Indicatieve Kaart van Archeologische Waarden 3e generatie*. Amersfoort: Rijksdienst voor het Cultureel Erfgoed.

English Heritage, 2015. The Historic Environment in Local Plans. Historic Environment Good Practice Advice in Planning 1, Swindon.

Flatman, J. and D. Perring, 2012. The National Planning Policy Framework and Archaeology: A Discussion. *Papers from the Institute of Archaeology* 22, 4–10. http://doi.org/10.5334/pia.390.

Fulford, M., 2015. The Towns of South-East England, in N. Holbrook and M. Fulford (eds), *The Towns of Roman Britain: the Contribution of Commercial Archaeology Since 1990*. Malet Street: Society for the Promotion of Roman Studies, 59-89.

Heaton, M., 2014. Constructing Archaeology: The Application of Construction Management Practices to Commercial Archaeology in Britain. *The Historic Environment: Policy & Practice* 5(3), 245–257.

Historic England, 2018. *Roman Settlements*. Introductions to Heritage Assets for Historic England, Swindon.

Historic England, 2020. *Mineral Extraction and Archaeology.* Historic England Advice Note 13, Swindon.

Holbrook, N, 2015. Introduction, in N. Holbrook and M. Fulford (eds), *The Towns of Roman Britain: the Contribution of Commercial Archaeology Since 1990*. Malet Street: Society for the Promotion of Roman Studies, 1-6.

Hopkins, D.W. 2004. Relevance of soil biology and fertility research to archaeological preservation by reburial, *Conservation and Management of Archaeological Sites* 6 (3-4), 167-175. https://doi.org/10.1179/135050304793137856.

Huggett, J., 2018. Reuse Remix Recycle: Repurposing Archaeological Digital Data. *Advances in Archaeological Practice* 6(2), 93-104. https://doi.org/10.1017/aap.2018.1.

Jackson, R. and G. Burleigh, 2018. *Dea Senuna: Treasure, cult and ritual at Ashwell, Hertfordshire*. London: British Museum Press.

Kamermans, H., J. Deeban, D. Hallewas, M. van Leusen, P. Verhagen and P. Zoetbrood, 2004. Deconstructing the crystal ball: The state of the art of predictive modelling for archaeological heritage management in the Netherlands, in K.F. Ausserer, W. Börner, M. Goriany and L. Karlhuber-Vöckl (eds), *Enter the Past: The E-way into the four dimensions of cultural heritage [CD-ROM]*. Oxford: Archaeopress (BAR International Series 1227), 1-25.

Kamermans, H., M. van Leusen and P. Verhagen, 2009. Archaeological prediction and risk management, in H. Kamermans, M. van Leusen and P. Verhagen (eds), *Archaeological Prediction and Risk Management: Alternatives to Current Practice*. Amsterdam: Amsterdam University Press (Archaeological Studies Leiden University 17), 9-18.

Kempf, M., 2019. The application of GIS and satellite imagery in archaeological land-use reconstruction: A predictive model? *Journal of Archaeological Science: Reports* 25, 116-128. https://doi.org/10.1016/j.jasrep.2019.03.035.

Kohler, T.A. and S.C. Parker, 1986. Predictive Models for Archaeological Resource Location. *Advances in Archaeological Method and Theory* 9, 397-452.

Kvamme, K.L., 1988. Development and testing of quantitative models, in W.J. Judge and L. Sebastian (eds), *Quantifying the Present and Predicting the Past: Theory, method and application of archaeological predictive modeling*. Washington: Government Printing Office, 325-348.

Lauwerier, R.C.G.M., M.C. Eerden, B.J. Groenewoudt, M.A. Lascaris, E. Rensink, B.I. Smit, B.P. Speleers and J. Van Doesburg, 2018. A Toolbox for Archaeological Heritage Management. Maps, Methods and More for Effective and Efficient Selection of Valuable Archaeology, *Internet Archaeology* 49. https://doi.org/10.11141/ia.49.8.

Lipovetsky, S., 2009. Pareto 80/20 law: derivation via random partitioning, *International Journal of Mathematical Education in Science and Technology* 40(2), 271-277, https://doi.org/10.1080/00207390802213609

Lockyear, K. and E. Shlasko, 2017. Under the Park. Recent Geophysical Surveys at Verulamium (St Albans, Hertfordshire, UK). *Archaeological Prospection* 24, 17-36. https://doi.org/10.1002/arp.1548.

Marwick, B. and S.E. Pilaar Birch, 2018. A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing. *Advances in Archaeological Practice* 6(2), 1-19. https://doi.org/10.1017/aap.2018.3.

Meffert, M., 2009. The high price or the first prize for the archaeological predictive model, in H. Kamermans, M. van Leusen and P. Verhagen (eds), *Archaeological Prediction and Risk Management: Alternatives to Current Practice*. Amsterdam: Amsterdam University Press (Archaeological Studies Leiden University 17), 33-40.

Menard, J.C., 2011. *Using GIS and Historical Data to Reconstruct the Ravenglass-Ambleside Roman Road, Cumbria, United Kingdom*. Minnesota (unpublished PhD thesis University of Minnesota).

Nakoinz, O., 2018. Models and Modelling in Archaeology. *Historical Social Research* 31, 101-112. https://doi.org/10.12759/hsr.suppl.31.2018.101-112.

North Hertfordshire Museum, 2019. *Roman Baldock: An exhibition at Baldock Museum (May 2018-May 2019)*. Museum brochure for the North Hertfordshire Museum, Hertfordshire.

Phillips, M., H. Duncan and C. Mallows, 2009. *Four Millennia of Human Activity Along the A505 Baldock Bypass, Hertfordshire.* Oxford: East Anglian Archaeology Report.

Podobnikar, T., T. Veljanovski, Z. Stanèiè and K. Oštir, 2001. Archaeological Predictive Modelling in Cultural Resource Management, in M. Konecny (ed), *GI in Europe: Integrative, Interoperable, Interactive (AGILE), Association of Geographic Information Laboratories in Europe, April 19-21 2001*. Brno: Czech Republic, 535-544.

Richards, J., 2017. Twenty Years Preserving Data: A View from the United Kingdom. *Advances in Archaeological Practice* 5(3), 227-237. https://doi.org/10.1017/aap.2017.11.

Rogers, A., 2013. *Water and Roman Urbanism: Towns, Waterscapes, Land Transformation and Experience in Roman Britain*. Leiden: Brill.

Secretary of State for Housing, Communities and Local Government, 2019. National Planning Policy Framework. Legislative Policy document presented to parliament by command of Her Majesty, Westminster.

Taylor, J., 2013. Encountering Romanitas: Characterising the Role of Agricultural Communities in Roman Britain. *Britannia 44*, 171-190. https://doi.org/10.1017/S0068113X13000202.

Tereszczuk, S., 2004. *North Hertfordshire and Stevenage Landscape Character Assessment.* Aylesbury: Babtie Group.

Van Leusen, M., 2002. *Pattern to process: methodological investigations into the formation and interpretation of spatial patterns in archaeological landscapes*. Groningen: University of Groningen Press.

Verbruggen, M., 2009. On costs and benefits in archaeological prospection, in H. Kamermans, M. van Leusen and P. Verhagen (eds), *Archaeological Prediction and Risk Management: Alternatives to Current Practice.* Amsterdam: Amsterdam University Press (Archaeological Studies Leiden University 17), 27-32.

Verhagen, P, H. Kamermans and M. van Leusen, 2009. The future of archaeological predictive modelling, in H. Kamermans, M. van Leusen and P. Verhagen (eds), *Archaeological Prediction and Risk Management: Alternatives to Current Practice*. Amsterdam: Amsterdam University Press (Archaeological Studies Leiden University 17), 19-26.

Verhagen, P. and T.G. Whitley, 2012. Integrating Archaeological Theory and Predictive Modeling: a Live Report from the Scene. *Journal of Archaeological of Method and Theory* 19, 49-100. https://doi.org/10.1007/s10816-011-9102-7.

Verhagen, P. and T.G. Whitley, 2020. Predictive Spatial Modelling, in M. Gillings, P. Hacıgüzeller and G. Lock (eds) *Archaeological Spatial Analysis: A Methodological Guide.* Oxfordshire: Routledge, 231-246.

Verhagen, P., 2009. Predictive models put to the test, in H. Kamermans, M. van Leusen and P. Verhagen (eds), *Archaeological Prediction and Risk Management: Alternatives to Current Practice*. Amsterdam: Amsterdam University Press (Archaeological Studies Leiden University 17), 71-122.

Verhagen, P., 2009. Testing archaeological predictive models: a rough guide, in H. Kamermans, M. van Leusen and P. Verhagen (eds), *Archaeological Prediction and Risk Management: Alternatives to Current Practice*. Amsterdam: Amsterdam University Press (Archaeological Studies Leiden University 17), 63-70.

Verhagen, P., H. Kamermans, M. van Leusen, J. Deeban, D. Hallewas and P. Zoetbrood, 2007. First thoughts on the incorporation of cultural variables into predictive modelling, in P. Verhagen (ed), *Case Studies in Archaeological Predictive Modeling*. Amsterdam: Leiden University Press (Archaeological Studies Leiden University Press 14), 203-213.

Verhagen, P., L. Nuninger, F. Tourneux, F. Bertoncello and K. Jeneson, 2014. Introducing the Human Factor in Predictive Modelling: a Work in Progress, in E. Earl, T. Sly, D. Wheatley, A. Chrysanthi, K. Papadopoulos, P. Murrieta-Flores and I. Romanowska (eds), *Archaeology in the Digital Era (CAA12): Computer Applications and Quantitative Methods in Archaeology, 26-29 March 2012*. Amsterdam: Amsterdam University Press, 379-388.

Verschoof-van der Vaart, W.B., K. Lambers, W. Kowalczyk, Q.P.J. Bourgeois, 2020. Combining Deep Learning and Location-Based Ranking for Large-Scale Archaeological Prospection of LiDAR Data from the Netherlands. *ISPRS International Journal of Geo-information* 9(5), 293. https://doi.org/10.3390/ijgi9050293.

Wheatley, D., 2004. Making space for an archaeology of place. *Internet Archaeology* 15. https://doi.org/10.11141/ia.15.10.

Wilcox, B., 2012. Archaeological Predictive Modelling Used For Cultural Heritage Management, in Z. Mingquan, I. Romanowska, Z. Wu, P. Xu and P. Verhagen (eds), *Revive the Past (CAA11): Computer Applications and Quantitative Methods in Archaeology, April 12-16 2011.* Amsterdam: Pallas Publications, 353-358.

Wilcox, W., 2014. Transparency, Testing and Standards for Archaeological Predictive Modelling, in E. Earl, T. Sly, D. Wheatley, A. Chrysanthi, K. Papadopoulos, P. Murrieta-Flores and I. Romanowska (eds), *Archaeology in the Digital Era (CAA12): Computer Applications and Quantitative Methods in Archaeology, 26-29 March 2012*. Amsterdam: Amsterdam University Press, 340–347. https://doi.org/10.1515/9789048519590-037.

Willems, W.J.H. and R.W. Brandt, 2004. *Dutch Archaeology Quality Standard*. Den Haag: Stichting Infrastructuur Kwaliteitsborging Bodembeheer.

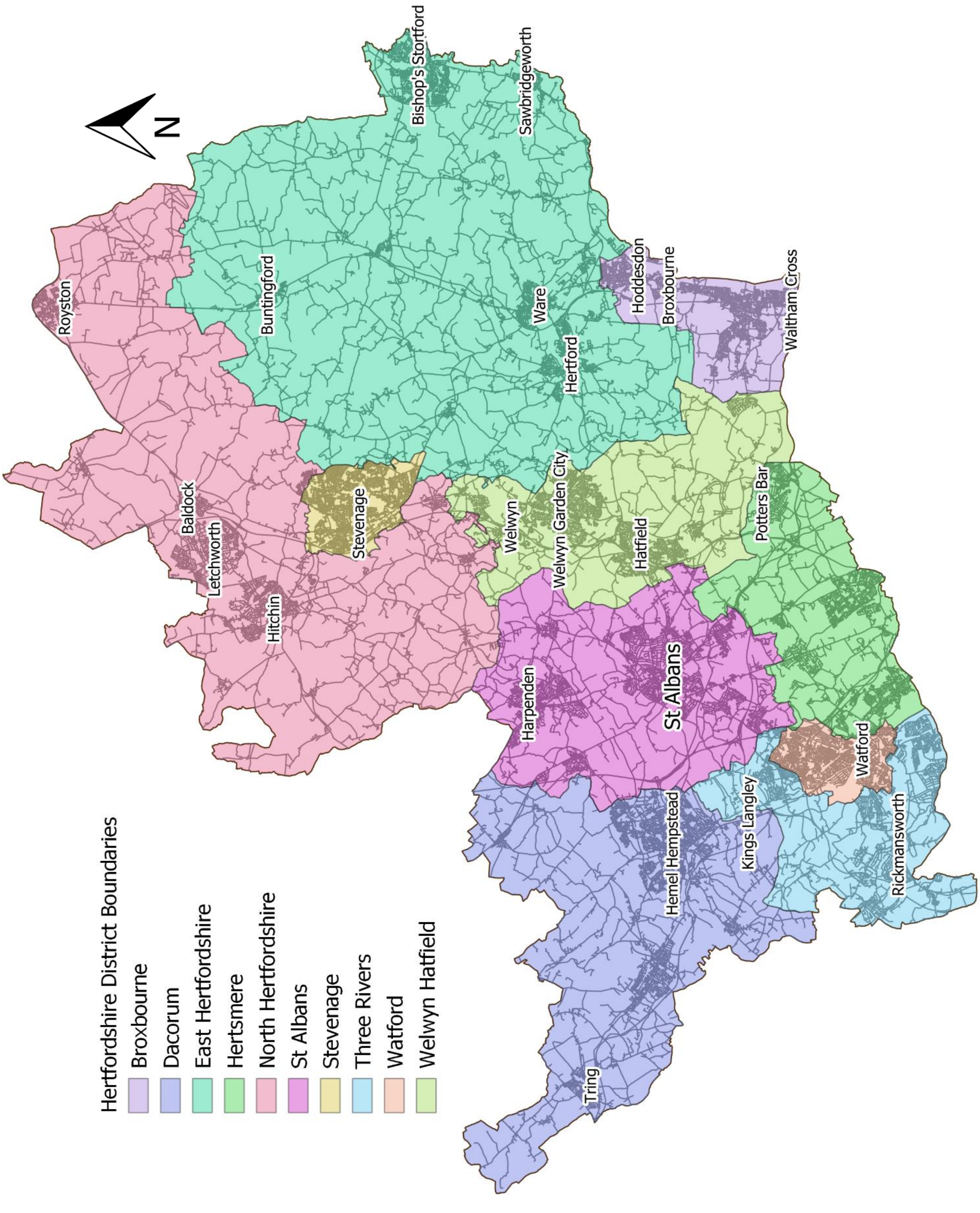
Wright, H. and J. Richards, 2018. Reflections on Collaborative Archaeology and Large-Scale Online Research Infrastructures, *Journal of Field Archaeology* 43(1), S60-S67 https://doi.org/10.1080/00934690.2018.1511960.

Figures

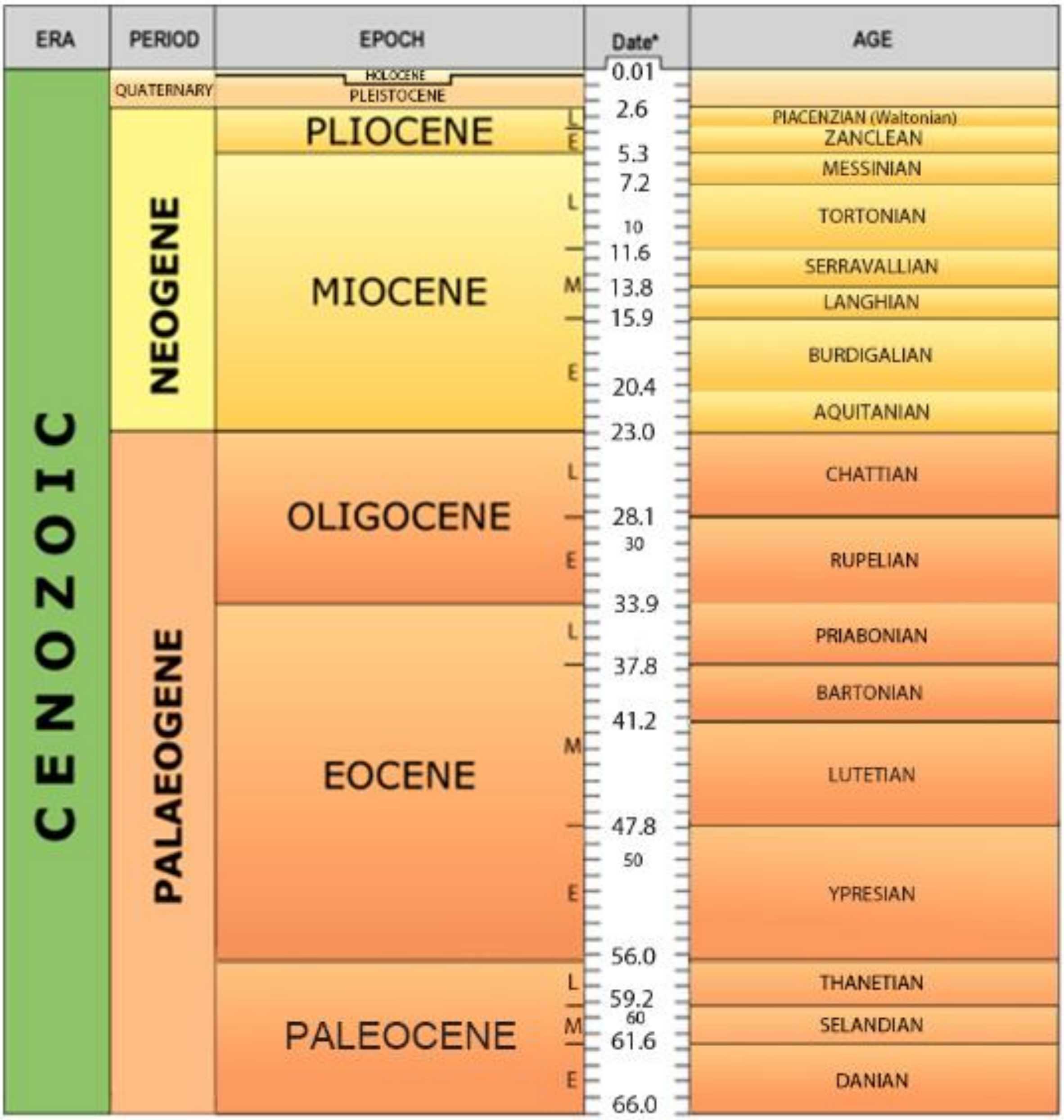
Tables

Appendices

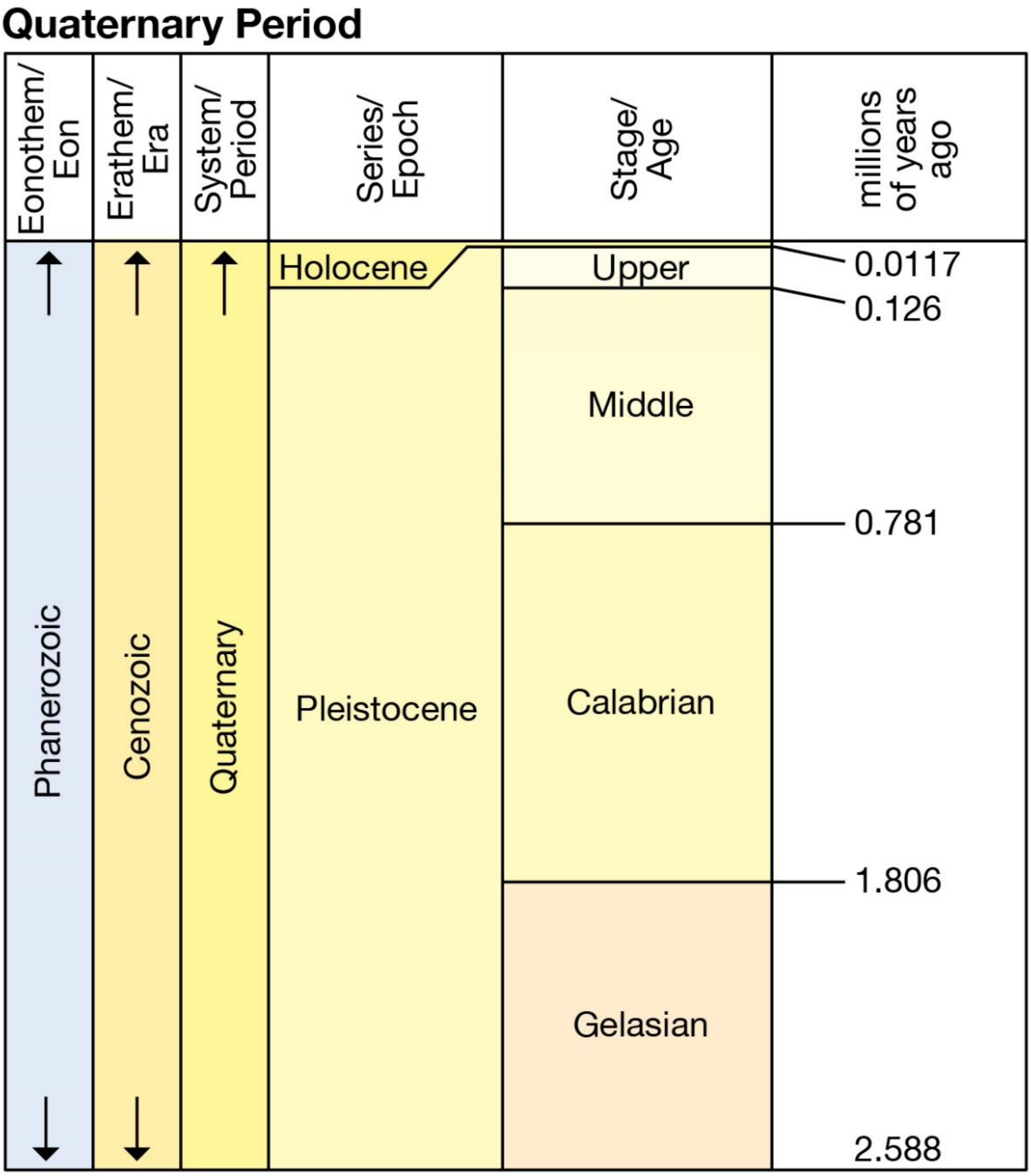
Appendix 1: The county of Hertfordshire split into its ten districts: North Hertfordshire, Stevenage, East Hertfordshire, Welwyn Hatfield, Broxbourne, St. Albans, Hertsmere, Watford, Three Rivers and Dacorum. Based upon the 'Local Authority Districts (December 2019) UK BFE' data source, with the permission of ONS Geography Open Data, and 'OS Open Roads', with the permission of Ordnance Survey.



Appendix 2: Timeline of the Cenozoic geological era (66 Ma – 0.01 Ma), from the Palaeogene till the Quaternary (www.bgs.ac.uk/discoveringGeology).



Appendix 3: Timeline of the Quaternary geological period, within the Cenozoic era (www.stratigraphy.org/timescale).



Published with permission from the International Commission on Stratigraphy (ICS). International chronostratigraphic units, ranks, names, and formal status are approved by the ICS and ratified by the International Union of Geological Sciences (IUGS).

Source: 2012 International Stratigraphic Chart produced by the ICS.

Appendix 4: Subjects categorised as settlement structures, within the 'Settlement' site type.

SETTLEMENT STRUCTURES	
Aisled building	Manor house
Basilica	*Municipium*
Bath house	Occupation site
Baths	Oppidum
Building	Outbuilding
Cellar	Outhouse
Country house	Palisaded settlement
Demolished building	Post built structure
Deserted settlement	Prison
Destroyed monument	Rectangular enclosure
Enclosed hut circle settlement	Round house (domestic)
Enclosed settlement	Ruined building
Extant building	Settlement
Forum	Shrunken village
Hall house	Structure
House	Temporary camp
House platform	Tessellated floor
Hut	Theatre
Hut circle	Town
Hut circle settlement	Villa
Hypocaust	Village

Appendix 5: Subjects categorised as settlement-related objects, within the 'Settlement' site type.

SETTLEMENT OBJECTS		
Amphora	Coin hoard	Phial
Amulet	Cup	Pillbox
Animal bone	Diadem	Pin
Animal remains	Dish	Plate
Artefact scatter	Figurine	Pot
Bead	Flagon	Pottery
Beaker	Gaming piece	Poun
Body sherds of vessel	Glass bowl	Ring
Bottle	Hearth	Roman deposits
Bowl	Jar	Scissors
Bracelet	Jug	Sherd
Brooch	Knife	Statue
Buckle	Lamp	*Strigil*
Cameo	Lamp holder	Toilet implement
Ceramics	Mortar (vessel)	Token
Clay pipe (smoking)	Needle	Vase
Clay pipe bowl	Oyster shell	Vessel
Coin	Personal ornament	

Appendix 6: Subjects categorised as infrastructure, within the 'Settlement' site type.

INFRASTRUCTURE	
Arch	*Mansio*
Artificial mound	Oval enclosure
Bank (earthwork)	Palaeochannel
Beacon	Pit
Boundary ditch	Pit alignment
Bridge	Pit cluster
Cess pit	Pits
Clay puddling pit	Pond barrow
Cobbled surface	Quay
Ditch	Rectilinear enclosure
Ditched enclosure	Ring ditch
Ditches	Road
Earthwork	Round barrow
Enclosure	Rubbish pit
Fire pit	Segmented ditch
Garden feature	Square enclosure
Garden path	Stock enclosure
Gate	Storage pit
Granary	Trackway
Gullies	Triumphal arch
Gully	Waster tip
Headland	Wharf
Hollow way	Wheel ruts
Landscape park	Yard
Linear earthwork	

Appendix 7: Subjects categorised as unknown settlement structure remains, within the 'Settlement' site type.

UNKNOWN SETTLEMENT STRUCTURE REMAINS	
Architectural component	Floor
Architectural fragment	Floor tile
Architectural fragments	Flue tile
Beam slot	Glass
Brick	Mosaic
Brick fragment	Nail
Brick wall footings	Posthole
Brick/tile	Roof tile
Bricks	*Tessera*
Building material	Tile
Ceramic building material	Tile burial
Column	Tile fragment
Daub	Wall
Flint foundations	Wall plaster

Appendix 8: Subjects categorised as agricultural processing sites, within the 'Economic' site type.

AGRICULTURAL PROCESSING SITES
Agricultural building
Corn drier
Corn drying kiln
Corn drying oven
Malt kiln
Maltings
Oven
Watermill
Windmill mound

Appendix 9: Subjects categorised as agricultural land, within the 'Economic' site type.

AGRICULTURAL LAND	
Aisled barn	Field
Ard marks	Field boundary
Barn	Field system
Coaxial field system	Livestock enclosure
Croft	Pasture
Drove road	Plough marks
Farm	Vineyard
Farmstead	

Appendix 10: Subjects categorised as industrial production sites, within the 'Economic' site type.

INDUSTRIAL PRODUCTION SITES	
Blacksmiths workshop	Lithic working site
Bronze working site	Mint
Chalk pit	Pottery kiln
Clay pit	Pottery manufacturing site
Dene hole	Quarry
Extractive pit	Quarry pit
Factory	Quarry pits
Furnace	Quern
Industrial site	Shop
Iron furnace	Smeltery
Iron working site	Tile kiln
Kiln	Workshop
Lime kiln	

Appendix 11: Subjects categorised as industrial objects, within the 'Economic' site type.

INDUSTRIAL OBJECTS		
4x residual pottery sherds	Flint scatter	Pick
Axe trimming flake	Graver	Scraper
Blade	Hammer	Seal matrix
Burnt flint	Hammerscale	Slag
Coin mould	Hammerstone	Spindle whorl
Core	Hoard	Steelyard
Crucible	Lithic implement	Trader's token
Debitage	Lithic scatter	Tranchet axehead
Flake	Loomweight	Worked flint
Flakes	Microburin	
Flint	Microlith	

Appendix 12: Subjects categorised as military structures, within the 'Military' site type.

MILITARY STRUCTURES	
Angle tower	Palisade
Curtain wall	Palisaded enclosure
Dyke (defence)	Postern
Fort	Rampart
Fortlet	Siegework
Gatehouse	Tower
Hillfort	Town defences
Moat	Univallate hillfort
Paddock	

Appendix 13: Subjects categorised as military objects, within the 'Military' site type.

MILITARY OBJECTS
Arrowhead
Axe
Helmet
Hipposandal
Hob nail
Horse trapping
Horseshoe
Leaf arrowhead
Spear
Weapon

Appendix 14: Subjects categorised as religious sites, within the 'Ritual' site type.

RELIGIOUS SITES
Ritual pit
Shrine
Temple

Appendix 15: Subjects categorised as funerary sites, within the 'Ritual' site type.

FUNERARY SITES	
Animal burial	Extended inhumation
Barrow	Grave
Barrow cemetery	Human bone
Bowl barrow	Human remains
Burial	Human skeleton
Burial pit	Inhumation
Casket burial	Inhumation burial
Cemetery	Inhumation cemetery
Cinerary urn	Long barrow
Cist	Mausoleum
Coffin	Mixed cemetery
Cremation	Mortuary enclosure
Cremation burial	Sarcophagus
Cremation cemetery	Tomb
Cremations	Urn
Crouched inhumation	Urns
Enclosed cremation cemetery	

Appendix 16: Subjects categorised as Roman water sources, within the 'Water sources' site type.

ROMAN WATER SOURCES
Arched brick culverts
Brick drain
Canal
Culvert
Drain
Drainage ditch
Ford
Leat
Pond
Sluice
Water channel
Water pipe
Watercourse
Well

Appendix 17: Subjects that were uncategorised due to them being too general, within the 'Miscellaneous' site type.

UNCATEGORISED
Cropmark
Feature
Find
Findspot
Layer
Linear feature
Site
Sub surface deposit
Tree hole
Unidentified object

Appendix 18: Subjects that were deemed to be unreliable observations due to them likely not dating to the Roman period, within the 'Miscellaneous' site type.

UNRELIABLE OBSERVATIONS
Adulterine castle
Air raid shelter
Barbican
Bastion
Brewery
Brewery buildings
Burh
Castle
Chapel
Church
Court of requests
Court room
Deer park
Grubenhaus
Leper hospital
Methodist chapel
Motte
Motte and bailey
Parish church
Ridge and furrow
Shell keep

Appendix 19: Comparison of the soil textures among researched and non-researched areas.

Soil of Researched Areas			Soil of Non-Researched Areas		
Soil Texture Groups	**Area (km²)**	**%**	**Soil Texture Groups**	**Area (km²)**	**%**
Loam Group	247.111	**66.4%**	Loam Group	872.503	**68.6%**
Silt Group	17.383	**4.7%**	Silt Group	89.893	**7.1%**
Sand Group	19.006	**5.1%**	Sand Group	55.350	**4.4%**
Clay Group	1.046	**0.3%**	Clay Group	15.349	**1.2%**
Mixed Group	87.543	**23.5%**	Mixed Group	238.574	**18.8%**
Total	372.089 km²	100%	Total	1271.669 km²	100%

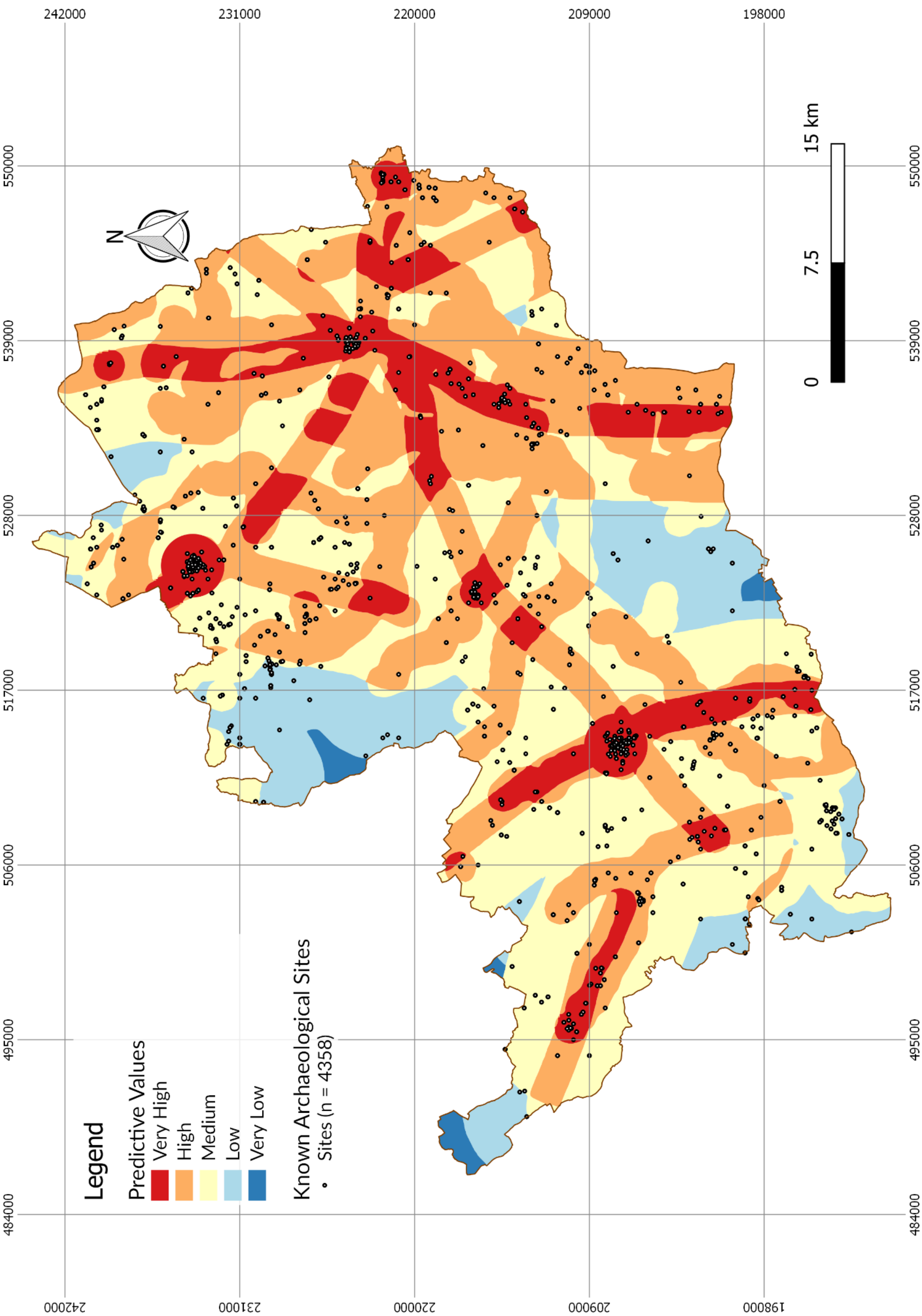
Appendix 20: Comparison of the groundwater levels among researched and non-researched areas.

Groundwater of Researched Areas			Groundwater of Non-Researched Areas		
Groundwater level	**Area (km²)**	**%**	**Groundwater level**	**Area (km²)**	**%**
Wet	290.199	**78.0%**	Wet	943.894	**74.2%**
Damp	3.113	**0.8%**	Damp	6.498	**0.5%**
Dry	78.777	**21.2%**	Dry	321.277	**25.3%**
Total	372.089 km²	100%	Total	1271.669 km²	100%

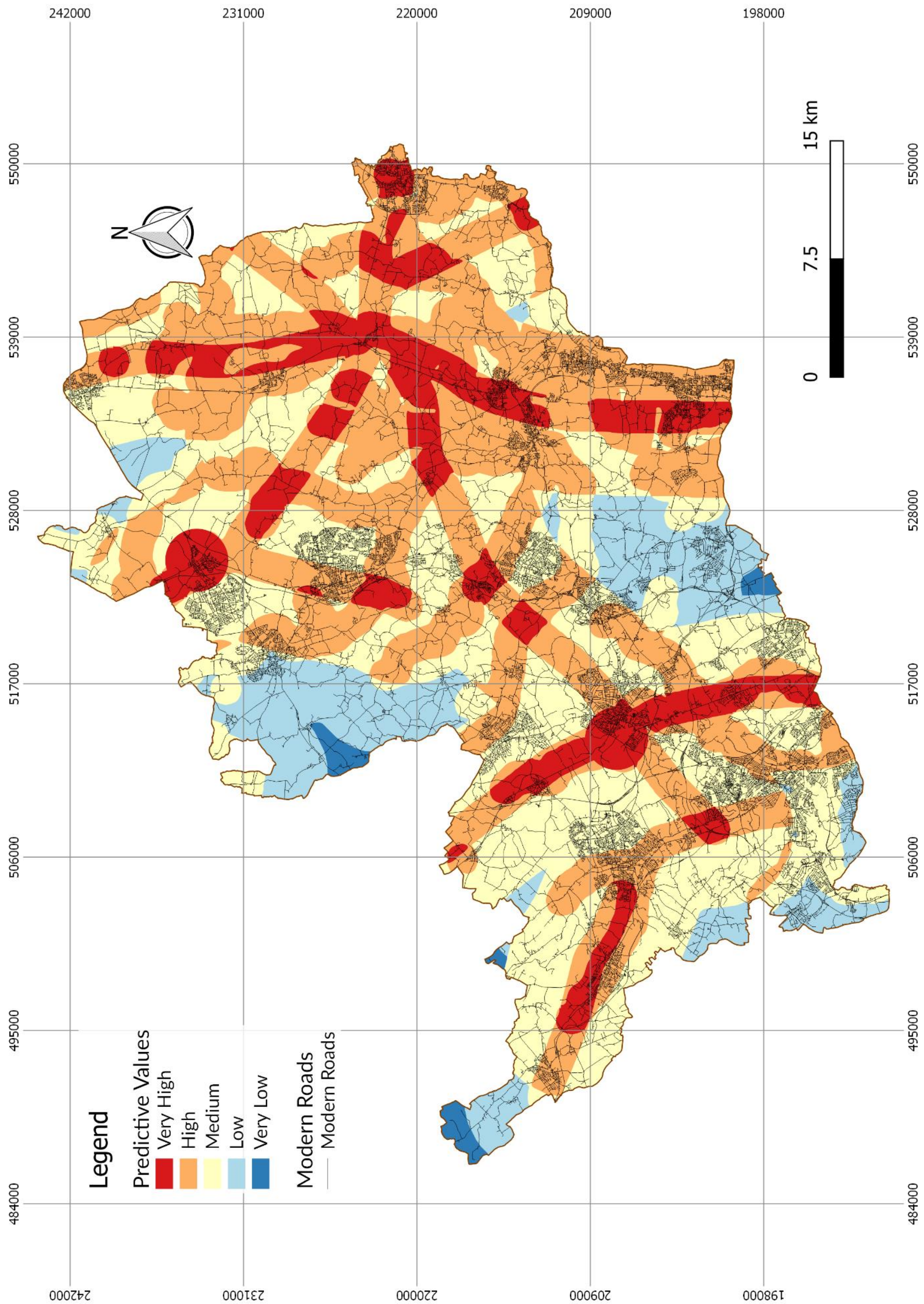
Appendix 21: Comparison of the land-use among researched and non-researched areas.

Modern Land Use of Researched Areas			Modern Land Use of Non-Researched Areas		
Modern Landuse	**Area (km²)**	**%**	**Modern Landuse**	**Area (km²)**	**%**
Urban area	131.582	**35.4%**	Urban area	220.840	**17.4%**
Cropland	218.809	**58.8%**	Cropland	970.314	**76.3%**
Forest and heathland	18.884	**5.1%**	Forest and heathland	75.623	**5.9%**
Roads and tracks	1.228	**0.3%**	Roads and tracks	1.565	**0.1%**
Waterbodies	1.496	**0.4%**	Waterbodies	3.020	**0.2%**
Total	372.089 km²	100%	Total	1271.669 km²	100%

Appendix 22: The Roman Hertfordshire predictive model, with Roman sites (n = 4358).



Appendix 23: The Roman Hertfordshire predictive model, with modern roads from the 'OS Open Roads' layer from permission of the Ordnance Survey.



Appendix 24: Advised intervention guide, based on the predictive values of the Roman Hertfordshire predictive model and the size of proposed development.

Predictive value	Size of Development	Advised Method of Intervention*
Very High	> 50 m^2	Geophysical survery of the development area. Core sampling and trial trenches within the development area +10% more in the wider area.
	< 50 m^2	No action
High	> 100 m^2	Core sampling and trial trenches within the development area +10% more in the wider area.
	< 100 m^2	No action
Medium	> 1,000 m^2	Coring sampling and trial trenches within the development area.
	< 1,000 m^2	No action
Low	> 25,000 m^2	Desk assessment and core sampling within the development area.
	< 25,000 m^2	No action
Very Low	> 50,000 m^2	Desk assessment of the development area.
	< 50,000 m^2	No action

*Advised methods of intervention are applicable to developments which will disturb the soil approximately 30cm or more beneath the surface.