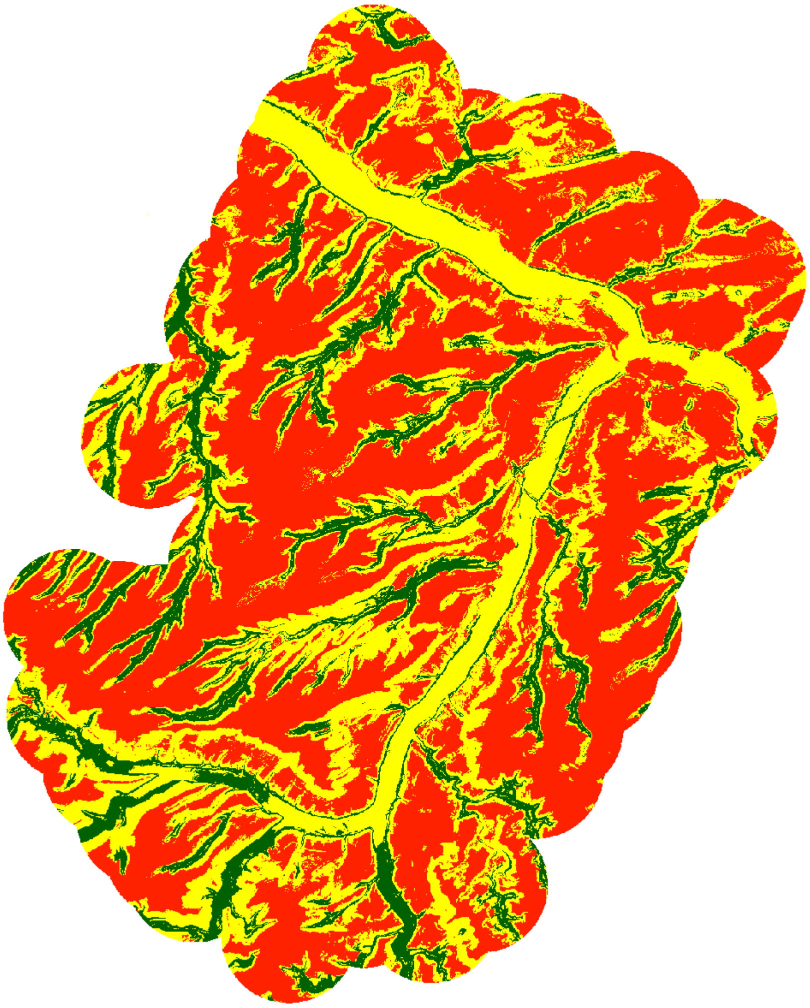
The Roman Settlement Pattern of the Somme:

Site Location analysis and Predictive Modelling

By

Nicolas P.A. Revert



Front page figure: Multivariate model of the physical landscape in region 2

The Roman Settlement Pattern of the Somme:
Site Location analysis and Predictive Modelling

Author: Nicolas P.A. Revert

Supervisor: Dr. K. Lambers

Specialisation: MSc Digital Archaeology

University of Leiden, Faculty of Archaeology

Leiden, 15th of June 2017: Final version

Table of contents

Acknowledgements

The completion of this thesis would never have been possible without many people who supported me and my work. I would like to give my absolute gratitude to Karsten Lambers, for having supervised my work with the most welcome diligence, kindness and for his exceptionally helpful critical eye. Without Philip Verhagen’s guidance and advices – within and besides the internship in Vrije Universiteit Amsterdam – I would never have been able to understand and implement predictive modelling in my thesis, and for this I am very grateful. I wish to express the utmost gratitude to Xavier Deru without whom this whole year of Digital Archaeology would never have been possible, and who accepted to share the data of the ABG program and kept a very kind and supporting interest towards my work and my international year.

I should also give my thanks to Antonin Nüsslein and Laure Nuninger for having accepted to share their work and tools to me. Javier Arce should also be thanked for having supported and recommended my candidature for this Master, and for having instilled in me the wish to study abroad. I wish to give my gratitude to Laëtitia Meurisse, who has very generously supported and advised me for five years already.

On a personal side, my greatest thanks go to my friend Ella, for having taken the time and patience to proof-read and comment my thesis. I wish to express my genuinely profound thanks to the best international friends I could hope for and who constantly supported me throughout the year: Daniel, Iita, Pierre-Yves, Lorna, Marie, Aida, Mar, Melissa and all those I have forgotten. My ‘Southern’ friends are far from being forgotten as they advised, supported and believed in me before and during this short but intense year: Sabine, Alexandra, Salocin, Becky, Lucie, André, Kathlèen, Jérome, Philippe, Guillaume, Alexandre, Anthony, with a very special mention to Anne-Sophie, who created the seed from which this wonderful year sprouted. I also wish to thank Luca and Leon, my fellow comrades of the Digital Archaeology specialisation.

Finally, for having constantly believed in me, despite the ups and downs of modern archaeology, for having supported and invested so much in me for so many years, I give my wholehearted thanks to my family, from Dublin to Alès and from Landerneau to Ch’Pas-d’Calais.

Chapter 1. Introductory chapters

1.1 General introduction

The Roman world was essentially a rural one, despite all the achievements of urbanism. The vast majority of people lived in the countryside, in a very diverse range of rural settlements, from the single-house farm to the complex village and the aristocratic villa. The distribution of these sites differs from region to region, as well as inside each region, which can be due to many possible factors. Through the choices made by the native inhabitants or the newcomers, the location of each settlement was carefully decided upon consideration of the physical environment as well as the human environment. Whereas agronomic sources do give us details on the potential factors involved in such choices, they only apply to an elite Mediterranean villa-landscape, concerned with the production of exportation goods such as olive oil or wine. Thus, there is little retrievable written evidence that could help from a better understanding of site location in the North-western provinces.

Northern Gaul, commonly named *Gallia Belgica*, is a place of contrasts and high diversity of rural occupation: at one end of the spectrum can be found palatial villas disseminated in the *civitas Treverorum* or the *civitas Suessonium* (the ancient territories around the towns of Trier and Soissons), while at the other end, the northern part of the *civitas Menapiorum* – around Cassel in Northern France – is devoid of stone-built rural settlement (Wightman 1985, 106–111). In the South-western margin of *Gallia Belgica* lies the area which will be studied throughout this thesis. Comprising four micro-regions, it is entirely contained within the modern department of the Somme, but represents parts of two *civitates*: the *civitas Ambianorum* – whose capital was Amiens – and the *civitas Viromanduorum* – whose capital was Saint-Quentin and then Vermand during the Late Roman period.

Widely explored through field survey and especially aerial survey, the Somme department is well-known for its staggering number of villas, at least 860 (Ben Redjeb *et al.* 2005, 196). These so-called villas generally take the form of large stone-built farms, with a residential complex of several rooms and functional buildings around at least two very large courtyards. This departs from compact plan of Mediterranean villas which often display a very large amount of rooms around a smaller central courtyard. Although the nature and appellation of these Gallo-Roman villas raises debate as to whether most of them should be called villas at all, taking them into consideration at the same time as the non-villa sites gives us the portrait of a very extensively settled area during the Roman period, comprising more than 2000 rural settlements in 6170 km^2, which amounts to a mean of 0.32 sites per km^2. Although the Roman period lasted from around 50 BC to the 5th century AD in the area, most of these rural settlements – when chronological data is available – are dated from the 1st century AD to the 4th century AD. Although the quantity of data in the Somme is very important, many research problems are still unsolved. We will now delve into this aspect.

1.2 Research problems and aims

Most researchers have not inquired in depth or detail the environmental or human factors playing a role in the distribution and evolution of sites in the Somme department. This type of inquiry is known as site location analysis, which is practiced in many scientific or commercial fields. So shunned is site location analysis of the Roman Somme that it is sometimes considered as illusory because of the seemingly uniform shape of occupation: 'tenter de corréler précisément les implantations antiques avec les contextes géographiques ou topographiques s'avère un exercice relativement ardu et illusoire. (...) nous pressentons une occupation assez uniforme sur la majeure partie du territoire et une mise en valeur de tous les contextes topographiques' (Ben Redjeb *et al.* 2005, 188). Other studies, even though they are concerned with regional datasets of Roman sites, do not devote much effort to explaining the origin of the settlement patterns they shed light on (Bayard *et al.* 2011). While it is certain that physical environmental differences are not as sharp in the Somme compared to other areas – such as mountainous Eastern France – its subtle variations might still play an important role, while the human environment may have the largest share of influence in site location.

Despite the lack of willingness to research this topic, a few attempts at site location analysis have been carried out in the area, but through the scope of a handful of excavated sites in very small areas, achieving only empirical claims (Blondiau 2014). During my precedent research master thesis in Lille University and its subsequent publication (Revert 2016; Revert (To be published)), my attempt at analysing the reasons behind the Late Roman site distribution in the *civitates Ambianorum*, *Morinorum* and *Bononiensium* (the western half of the Somme and the Pas-de-Calais) was unsatisfactory, as the dataset of 290 sites – spread over 8316 km^2 – barely enabled any sort of understanding other than a slightly higher proportion of sites nearby rivers and Roman roads. Moreover, the evolution of site location through time was surprisingly seamless. Although the methodology of the analysis in itself is strongly at issue, the scarcity and the lack of representation of settlements during the Late Roman period should be the prime factor in these results, thus tending to corroborate the 'illusory' aspect of the analysis.

The archaeological research in the Somme is not as homogeneous as one might think: the plateaus were the main focus of aerial and field surveys, for they preserve artefacts and structures much better than silted-up valleys. The large-scale excavations were also mainly carried out in the plateaus, because this topographical category is the most suitable for large construction projects like highways and vast economic or industrial development zones. Despite these biases, many smaller urban and rural developments have led to the discovery of numerous sites in the valleys and the alluvial plains.

In order to deal with the complexity of a large dataset, new methods of analysis are therefore required. If it is arduous to solve problems of lack of representation without carrying out very large survey, coring and trenching campaigns, a better understanding of site location during the Roman period may still achieved. For this purpose will be designed and implemented an analysis method which is strongly based on predictive modelling.

Predictive modelling – although its original purpose was indeed to predict sites or other elements in diverse scientific fields – can also constitute a very useful tool towards building and testing different variables which may improve or impede the probability of site location. This in turn leads to several possible applications and objectives, such as the extrapolation of site location probabilities for heritage management or development planning. Though, in this thesis, the only use of this tool will be to assert the influence of a set of variables towards site location, which is known as Archaeological Location Modelling (ALM), a sub-type of predictive modelling.

This methodology will enable us to create models of site probability with weighted and combined variables (weighted multivariate models), whose correlation with the actual archaeological data will give us the opportunity to discuss their relevance in site location analysis. Models will be built quasi-completely independently of the archaeological dataset, but can only be statistically evaluated and commented upon with the help of the latter. More details on the specifics of predictive modelling and the methodology designed for this thesis will be given below, but first shall now be better defined the scope of this study.

1.3 Geographical area

1.3.1 Micro-regional approach

As studying the entire department of the Somme as one homogeneous ensemble would result in creating a very large and heterogeneous dataset – therefore considerably diminishing the effectiveness and relevance of the models – micro-regions have to be defined. For this purpose, a set of criteria forms the basis upon which decisions are made.

First and foremost, each micro-region must display a large enough number of sites (at least one hundred). Without this precondition, any statistical evaluation of the models' behaviour would be futile, as proportions feature heavily in the present method and these cannot be accurate below a hundred total elements. The Somme valley constitutes a very important area in terms of site frequency, on its entire course, but the neighbouring plateaus also display important site densities.

Secondly, there must be a satisfactory number of sites which are excavated well enough so that a chronological assessment of the models would be applicable. This means that areas with many recent excavations should be preferred to areas where there are comparatively larger amounts of surveyed sites, but which do not give any chronological information. The micro-regions must therefore aim at including the large scale projects such as highway, airport or canal constructions.

Thirdly, the micro-regions must share at least one similarity which justifies the modelling approach, for the models will be created on a very large scale and evaluated at a micro-regional level. The Somme River constitutes this common element, which runs East to West through the entire department. All the micro-regions will therefore include a part of the Somme River.

But a complete homogeneity between regions is also to be discarded; otherwise it would become impossible to assert the different settlement patterns in the Somme. The final criteria for the definition of the micro-regions are therefore their local specificities. These include the *civitas* which the micro-regions are part of, where we be included a micro-region which is entirely located in the *civitas viromanduorum*, while the other micro-regions are in the *civitas ambianorum*, meaning that some settlements would rather revolve around Saint-Quentin or Vermand than around Amiens. In terms of physical environment, the main tributaries of the Somme have very distinct characters which influenced local landscape. It is therefore desirable to include micro-regions which also contain large stretches of major rivers. The presence or lack of meanders in the Somme valley could also constitute diverging factors in site location and the Middle Somme area should therefore be isolated for its specificity. Finally, the relative ruggedness of each area differs widely, although never reaching the level encountered in more contrasted regions such as in Eastern or Southern France. The Santerre is a vast natural region of the Somme – in its Eastern reaches – which displays a generally very low ruggedness, with largely flat expanses of fertile and light soils, whereas the Amiénois – around Amiens – and the Vimeu further West display more frequent dry valleys and wider elevation differences.

1.3.2 Geography of the micro-regions

The four micro-regions that will be studied are therefore all part of the Somme department, in the newly created Hauts-de-France region, the fusion of the Picardy region with the Nord-Pas-de-Calais region (Fig. 1). The Somme, with its 6170km^2 extent, is still a very rural area, containing only one important town: its regional capital, Amiens. The Somme River is the most important element of the department’s landscape. It takes its source eastwards – in the neighbouring Aisne department – and follows a westwards and gently grading descent on 192km before reaching the Channel (Ben Redjeb 2013, 72–76).

The Somme River’s course can be divided into three sections: the Upper Somme, the Middle Somme and the Lower Somme, which are respectively included in the natural regions of the Santerre, the Amiénois and the Vimeu or Ponthieu. The four micro-regions are spread upon the length of the Somme River, while also taking into consideration the plateaus and the tributary rivers. The Upper Somme valley is flat and marshy, following a south-east to north-west direction before going northwards. Starting at Péronne, the Middle Somme valley follows a westwards course. Its small meanders are deeply cut into the plateaus, down to 60m of depth and sometimes displaying a 40% slope gradient (Ben Redjeb 2013, 75). From Amiens downwards, the Lower Somme is a marshy, flat and wide valley (1200m on average). It cuts the plateaus on a depth of up to 90m, in relatively gentle terraces. Alluvial silts and peats are deposited in the major river valleys: the Somme, the Avre, the Noye, the Hallue, the Ancre and the Airaine.

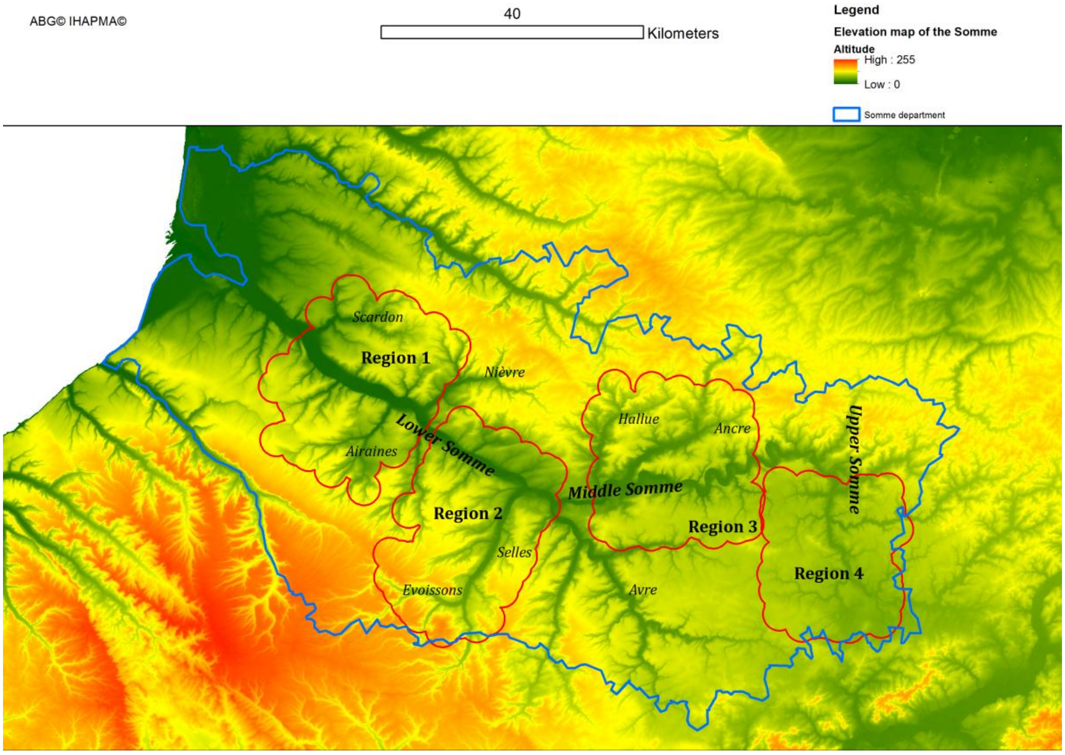


Figure 1: The four studied micro-regions (red), defined by a buffer area of 2 500m around the sites, ABG©

In general, the area is characterised by a vast plateau, culminating at 215m in the south-west and cut by the Somme River as well as its 26 tributaries, generally running south-west to north-east on the southern bank of the Somme or north-east to south-west on the northern bank. Three important rivers do not follow this rule: the Avre – the main tributary of the Somme, running south-east to north-west until the confluence at Amiens – as well as the Bresle and the Authie, running from the south-east into the sea and constituting the south-western and north-western limits of the department. A vast number of dry valleys are spread all over the plateaus, more abundant than the wet valleys, especially in the Amiénois. Their slopes can be as steep as 30%, but are generally quite gentle.

The geological strata of the plateaus are composed of a Turonian limestone base, over which are found Senonian flint clays, whose outcrops are often present on the borders of the plateau and on top of which often resides a layer of loess whose thickness varies from 1m to 10m. It is thickest on the upper central plateaus and especially in the easternmost part of the department, between the Upper Somme and the Avre, while it thins down towards the north-west. It is occasionally present on top of small pockets of sand. The erosion of the plateaus resulted in silt colluviums on north and east-facing slopes while west- and south-facing slopes are covered by more chalky and flinty colluviums. Mainly the maritime plain – which will not concern us further – was substantially changed and transformed during the Holocene.

Now will be detailed the local characteristics which justify the choice of the four micro-regions, which are mainly concerned with functional needs as criticised by Wheatley (1992, 136-137), but also some socio-cultural aspects.

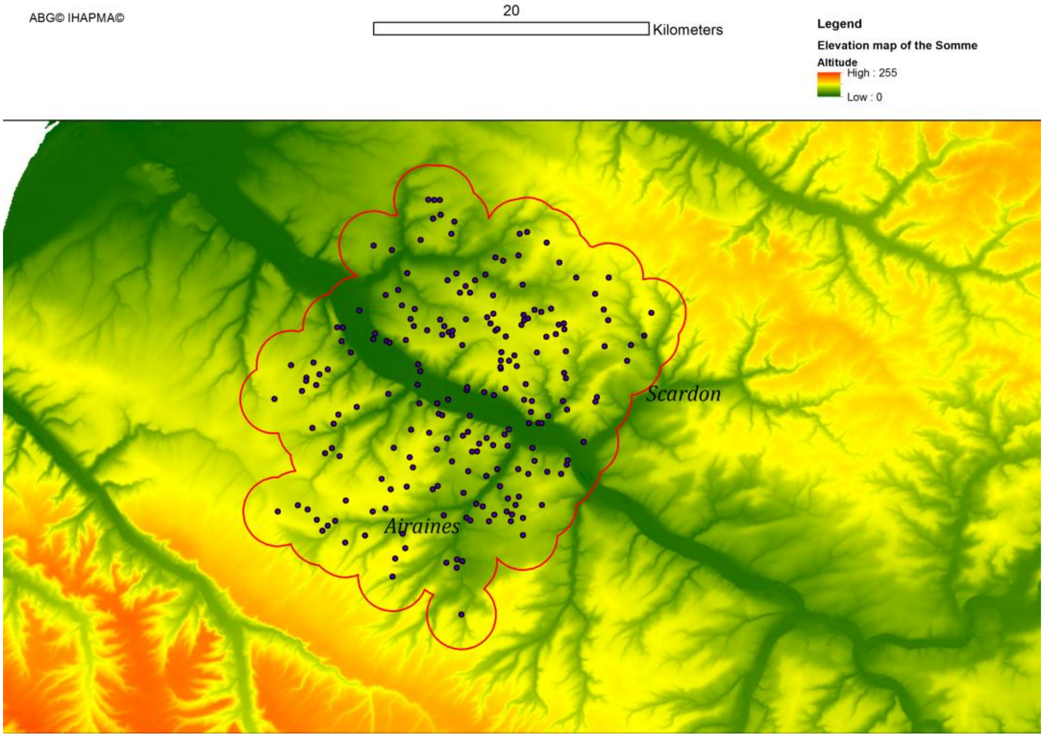


Figure 2: Micro-region 1 with its rural settlements, ABG©

-Micro-region 1: the Lower Somme (Figure 12)

The first micro-region is the westernmost, including a large part of the Lower Somme. Its 584km^2 extent comprises neighbouring plateaus as well as relatively minor tributary rivers: the Airaines in the South-east and the Scardon in the North-east. The topography is relatively low here in comparison to the other three micro-regions, but the Lower Somme cuts deep valley slopes, especially on its left bank. The Southern part of the region is included in the Vimeu natural region, while the Northern part is included in the Ponthieu. These two natural regions are virtually identical in terms of environmental characteristics, which are much gentler than in micro-regions 2 and 3, but still more rugged than micro-region 4 in the Eastern Santerre. The area was mainly excavated before highway constructions.

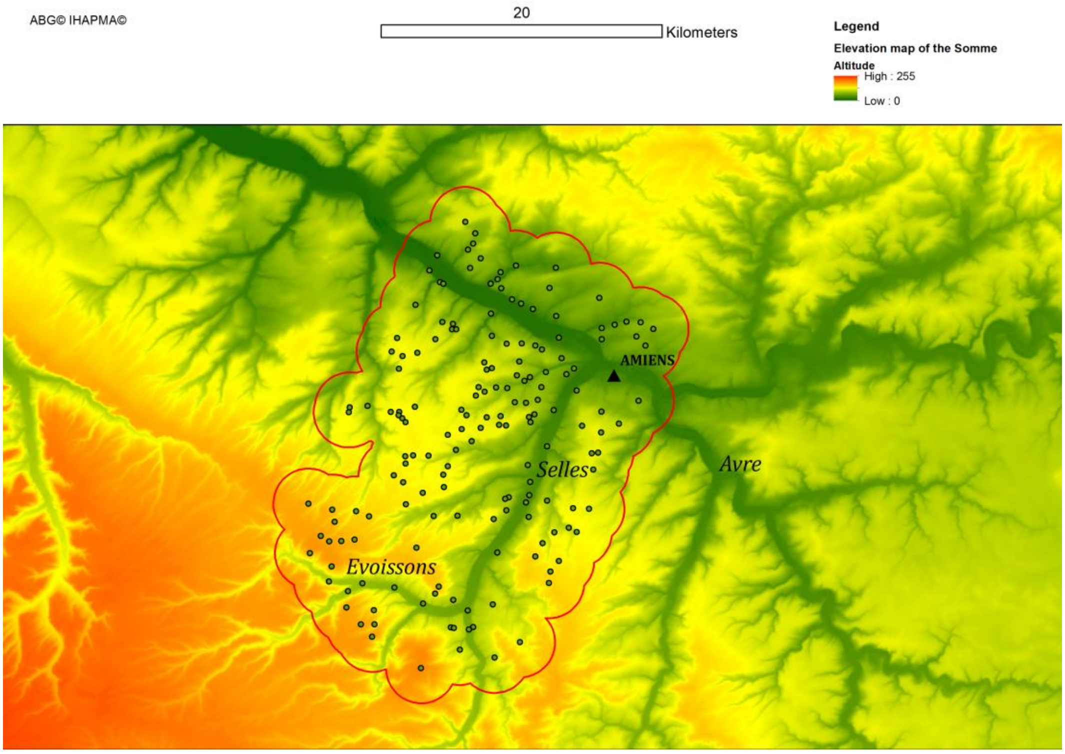


Figure 3: Micro-region 2 and its rural settlements, ABG©

-Micro-region 2: the Selle (Fig.3)

This area includes Amiens as well as a small part of the Lower Somme valley. Its plateaus and valleys are part of the Amiénois, mainly following the course of the Selle, a large tributary of the Somme running from the south-west and its main affluent: the Evoissons. Spanning 555km^2, this micro-region is characterised by a more contrasted topography, with steep slopes in its southern part near the Selle and the Evoissons, but elsewhere low-lying marshes, valleys and plateaus, especially near Amiens, where the Avre and the Selle meet the Somme. The very close proximity of Amiens and the importance of the Selles are the main specificities of this micro-region. Its excavations preceded urban developments in the vicinity of Amiens and highway constructions.

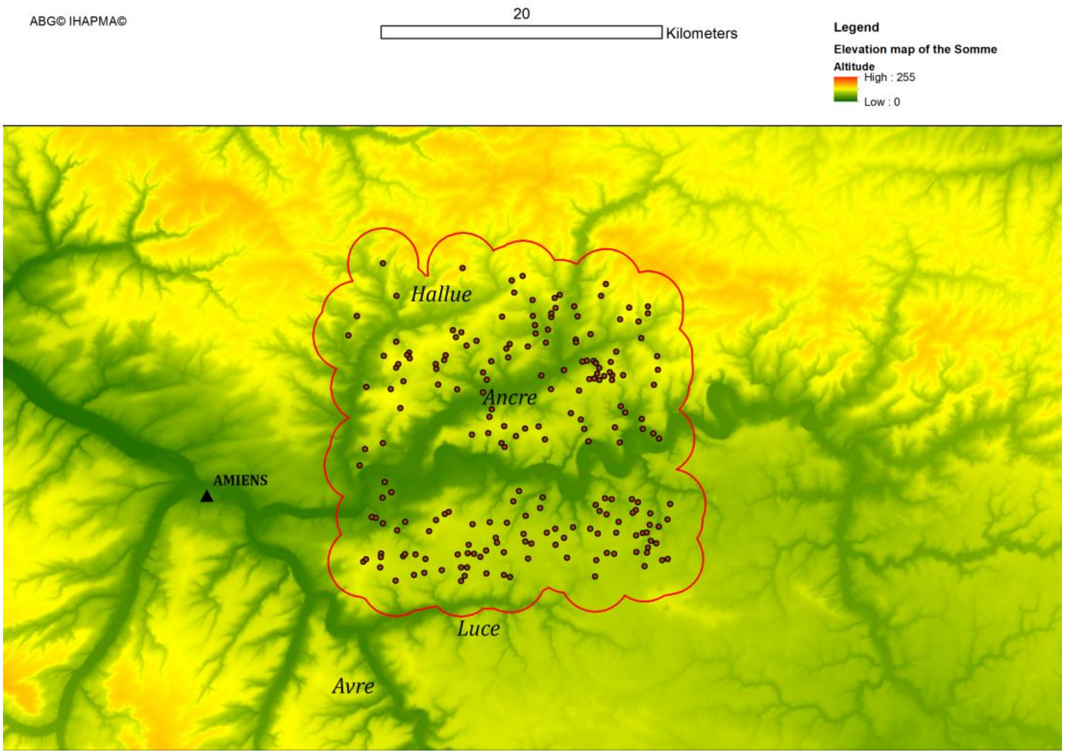


Figure 4: Micro-region 3 and its rural settlements, ABG©

-Micro-region 3: the Middle Somme (Fig.4)

The third micro-region includes the easternmost part of the Middle Somme and its meanders, as well as a large part of the Ancre and Hallue tributary rivers. Even though it is part of the Western Santerre, the slopes are relatively steep in this entire area, especially on the banks of the meanders. The 522km^2 area also includes a large part of the southern plateau up to the Luce basin, a tributary of the Avre. This area's specificities are therefore its meanders and its omnipresent hydrographical elements, involved in carving small stretches of plateaus. Its excavated settlements were discovered before highway constructions.

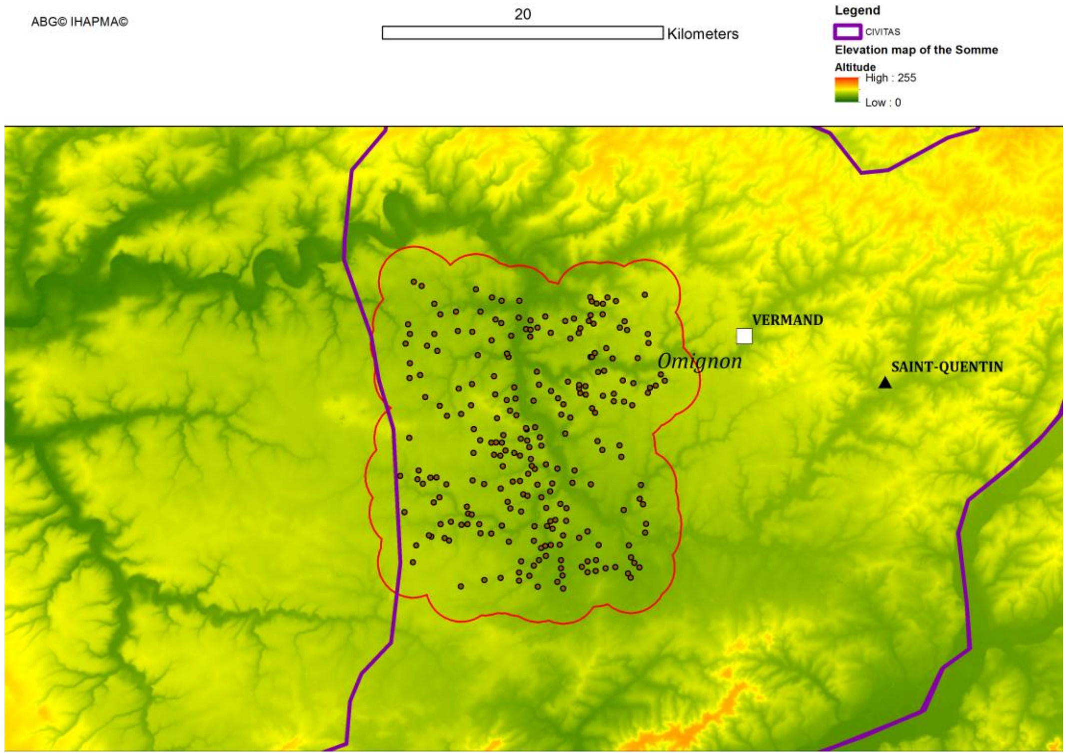


Figure 5: Micro-region 4 with its rural settlements and the approximate frontiers of the *civitas Viromanduorum*, ABG**©**

-Micro-region 4: the Upper Somme (Fig.5)

This micro-region is the most distinct, as it is characterised by a slightly drier climate, much thicker loess deposits and marshy valleys. It is also the only one that was not part of the *civitas Ambianorum* but rather the *civitas Viromanduorum*, at the eastern limit of the modern department. Slightly smaller than the other three areas (399km^2), this one comprises a very flat plateau which is part of the Santerre natural region, only cut by the marshy Upper Somme, whose valley floor is quite shallow. Some minor tributaries draw shallow valleys, but the Omignon valley is slightly steeper. Slightly to the east lies Saint-Quentin in the Aisne department, the ancient capital city of the *civitas Viromanduorum* and Vermand, which replaced Saint-Quentin as capital in the Late Roman period. This area is therefore distinct in terms of physical environment and cultural affiliation, but also in terms of excavations, as there were many highway constructions and the very large Canal Seine-Nord-Europe. Now that the general geographical setting of subject has been defined, some effort will be taken to set and explain the chronological boundaries of this study.

1.4 Chronological boundaries

Regarding the chronological scope of this study, it seemed unfeasible to include pre-Roman and post-Roman settlements, as this data is much less represented and not available in this dataset. The same applies to 'transitional' settlements which span from the Late La Tène period to the Gallo-Roman period, i.e. from the second half of the 1st century BC to the first half of the 1st century AD. Their exclusion from this study is due to them still being poorly understood on a global scale, and they also seem to vary widely in terms of evolution creation date – which can be very recent as well as in the Late Iron Age – and abandonment date, the latter event often occurring during the 1st century AD, while the remainder radically change during the same century.

The farms of this period have been the focus of scholarly debates for many decades now, as the regional variations, both in morphology, size and chronology, forbid any definitive conclusion (Collart 1996; Bayard and Collart 1996). Indeed, it comes as particularly debilitating for a chronological study that these farms/*fermes indigènes*/*fermes gallo-romaines précoces* are virtually indistinguishable from earlier La Tène farms and sometimes only display a slight displacement of the residential space (Ben Redjeb 2013, 103; Ben Redjeb *et al.* 2005, 191).

Considering the end of this study's timeframe, the hiatus separating the Late Roman period from the Merovingian period is not understood very well at all. The near entirety of settlements indeed display abandonments during the period from the end of the 3rd century AD to the beginning of the 5th century AD, while some disappear as early as the end of the 2nd century AD (Revert 2016a, 2016b). Only in the end of the 5th century AD or later are a few rural settlements again appearing in the archaeological record. While a huge depopulation of Northern Gaul is certain, the erosion of Late Roman strata and the very humble nature of finds also play a role in this unsolved hiatus. As far as the archaeological remains can indicate, only the *caput civitatis* of Amiens seemed to be continuously occupied, albeit in very uncertain modalities. The chronological boundaries of this study are therefore set during the Roman Imperial period: from the second half of the 1st century AD to the end of the 4th century.

As lengthy as this 350 years study period may seem like, it is possible to study it in the four micro-regions without having to segregate different groups of rural settlements. Indeed, the physical landscape does not seem to have changed much during this time-lapse. From the late Iron Age onwards, the landscape of Gallia Belgica was extensively open, be it in Champagne or in the heavier soils of Picardy (Wightman 1985, 4). After the 2nd century BC, the climate was also warmer and drier than before and although the Late Roman period and Early Middle Ages may display a slightly more important influence of rains, the climate was globally analogous (Wightman 1985, 5).

The same cannot be said of the coastal fringe – which is out of the present scope – displaying important rises in maritime influence, mainly due to strong storms breaking the thin sand dune barriers protecting the very flat expanse of marshy, peaty or immerged lands of the maritime plain, in a pattern similar to what is now recognised episodically in the Flemish maritime plain as well (Gandouin *et al.* 2007, 27; Ervynck *et al.* 1999, 101–107; Briquet 1930, 396; Meurisse *et al.* 2005, 683; Sommé 1977, 545–547). No such radical change in the landscape of the hinterland can be seen through the very limited geomorphological information available in the inland Somme. At most, the region may display – as in the Flemish maritime plain – alluvium deposits in the lower valleys and the borders of the maritime plain (Ervynck *et al.* 1999, 101).

Furthermore, the few palaeoenvironmental studies such as carpology, palynology and anthracology emphasize continuity rather than difference between the Early Roman period and the Late Roman period. Only a few changes occur in terms of proportions of hardy cereal species being cropped more frequently than white wheat, but both hint at the same agricultural traditions and the open field cultures they entail (Lepetz *et al.* 2003, 28; Matterne 2001; Revert 2016, 104). It is as yet unknown if the Somme region manifested a Late Roman resurgence of forests, as can be seen more to the South and the East, in the Vosges, in Lorraine, Argonne, in the forests of the Haye, Châtillon-sur-Seine and especially in the Ardennes (Wightman 1985, 263; Ouzoulias 2014).

Finally and most importantly, the justification for these chronological boundaries depends on the dated settlements of the dataset, which display a complete continuity and legacy from the Early Roman Empire to the Late Empire: around 300 AD, only four sites were created later than the period 100-150 AD. Settlement patterns and site location are therefore dependent on the Early Roman setting; the only changes occurring – as will be detailed below – concern the disaggregation of this setting rather than the creation of a new one.

The chronological boundaries having been defined and justified, the dataset of this thesis will now be presented.

1.5. Dataset

1.5.1 Rural settlements

All of the following data is part and property of the ABG program (Atlas des provinces romaines de Belgique et de Germanie), whose director is X. Deru (Université de Lille SHS) and to which I have contributed during the completion of my RMA in Gallo-Roman archaeology in the University of Lille.

The 822 rural settlements of this dataset provide the most representative element of the landscape, displaying morphological and chronological homogeneity and therefore following similar criteria for site location during the studied period, as shown by the graph below which indicates the entire chronological information retrievable from the 335 rural settlements which possess any (fig.6).

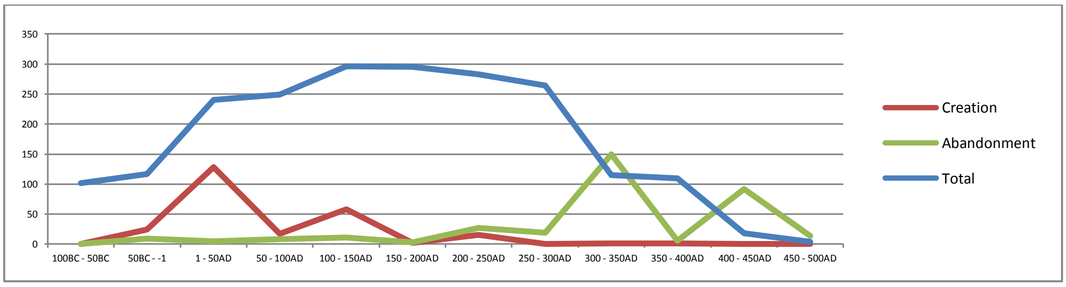


Figure 6: The evolution of the number of dated rural settlements, their creation and abandonment during time, ABG©

It should be noted that the second half of any century is much less representative than the first, as only very few sites possess chronological dating to the precision of half a century, but rather of an entire century. Most creations attributed to a century in general are therefore arbitrarily placed in the first half of this century's chronological classification.

Notwithstanding an ideal trend which would be much smoother, it should be clearly apparent that site creation occurs very early. When this study's chronological frame starts, in 50 AD, most sites are inherited from the first half of the 1st century and will remain so until the end boundary in 400 AD. The chronological classification applied here is very simplistic, allocating the maximal time span according to the termini post quem and ante quem provided by the description of survey data or excavated sites, which differs widely from period to period, as some materials are more precisely dated than others. The provision of a more recent dataset, with potentially more precise chronological evidence, other classification methods can be employed, which will be discussed in Chapter 4.

First of all must be explained how a chosen settlement can be defined as a villa or a ‘simple’ farm. Thanks to Roger Agache and other aerial surveyors, ‘villa studies’ have been very active in the Somme (Agache 1978; Agache 1975; Agache 1975; Agache *et al.* 1965; Bayard and Collart 1996; Bayard 1993; Bayard *et al.* 1996; Collart 1996; Pannetier 1996). The ‘traditional’ definition of what a villa is in the north of Gaul has therefore heavily been influenced by these early works, for which a villa is ‘an homogeneous set of substructure ordained around one or two courtyards, which serves as housing as well as production functions, storages: differently put, an ensemble which evokes a large farm’ (Agache 1978, 281). Other researchers put more emphasis on the architectural aspects of the villa which should be characterised by ‘a search for monumentality and a rigorous and symmetrical spatial organisation’ (Bayard *et al.* 1996, 5). One representative example of what these definitions represent can be found in Dury, where the residential building is of the most commonly found type: located at one end of the villa in a specific court, it possesses two aisles on the smaller sides and a covered portico on the front (fig.7).

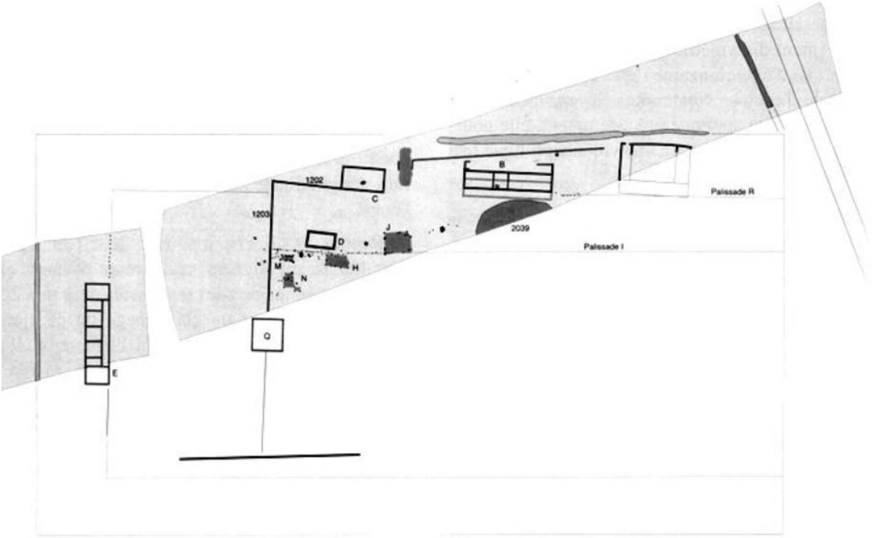


Figure 7: The so-called villa of the ‘Camp Rolland’ in Dury (Quérel *et al.* 2000, 38)

Recent studies on villas in the North-west of Europe have shown that this classification is erroneous, for villas are also defined as farms which are an integral part of the socio-economic Roman world (Roymans *et al.* 2011, 2; Habermehl 2013; Habermehl 2011, 62; Ferdière *et al.* 2010). Using well-excavated examples, Roymans and Derks have defined three categories of rural settlements in the north of Europe (Roymans *et al.* 2011, 2): villas, stone-built isolated farms and 'traditional' farms exclusively built of wood and earth. According to this classification, villas are large or medium stone-built farms with multiple production buildings and a residential complex possessing at least 15 rooms, generally with a floor- and/or wall-heating system (hypocausts), for which a good example can be found in Estrées-sur-Noye (fig.8).

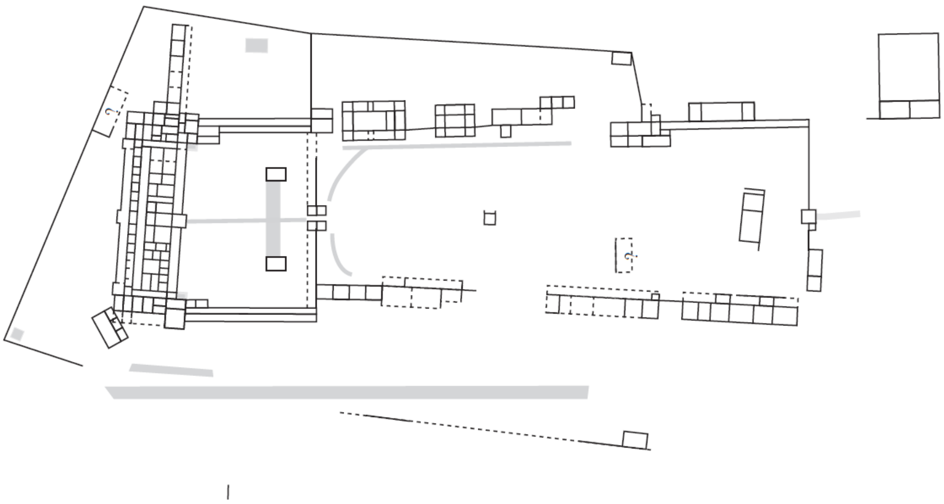


Figure 8: The very large villa of the 'Bois des Célestins' in Estées-sur-Noye (Ferdière *et al.* 2010, 420)

Stone-built farms therefore should pertain to all other stone-built settlements – including Dury – which do not fit this restricted definition of the villa. This classification is in fact very accurate, in the sense that there can be no more doubt as to the hierarchy of settlements fitting this definition of villas, and this represents more faithfully what Roman authors allude to when using the word villa: the center of a large rural domain/*fundus*, including this monumental centre as well as the *ager*/land (Wightman 1985, 105). This scholarly debate is not solved at all, as villas indeed share similarities but do differ widely in every region of the empire. If many more sub-classifications exist and can be applied in accordance with Roymans' and Derk's definitions, for this modelling purpose will be kept this tripartite division of rural settlements, adapted to the dataset of the Somme and its description in the ABG program, detailed in table 1 below.

The following paragraphs will now detail the origin of rural settlements in each micro-region, whose composition is summarised below, at the same time as the material criteria for each type of settlement (table 2).

Rural settlements	**Total**	**Villas**	**Stone-built farms**	**Post-built farms**
Criteria		**Large farm of at least 2ha**	**Small or large farm**	**Small or large farm**
		Mainly stone build structures	**At least one building with stone foundations**	**No evidence for stone foundations**
		Residential building with at least 15 rooms	Optionally roof tiles	
		Production buildings around a separate courtyard	Optionally evidence for heating system	
		Evidence for heating system		
		Optionally mosaics		
Micro-region 1	208	7	182	19
Micro-region 2	171	6	150	15
Micro-region 3	204	12	177	15
Micro-region 4	239	7	220	12

Table 1: Morphological criteria distinguishing types of rural settlements and hierarchical composition of the rural settlements in each micro-region, after Roymans *et al.* 2011, 2 and Deru 2012. ABG© IHAPMA©

-Micro-region 1: This area's 17 partially or fully excavated sites mainly result from the construction of the A16 highway on the right bank of the Somme.

-Micro-region 2: This region includes 35 excavated sites, which mainly stem from the development of the larger agglomeration of Amiens, as well as the A29 and A16 highways, respectively in the South-west and the North-west.

-Micro-region 3: The 18 excavated settlements of this area mostly result from the construction of an industrial airport facility around Méaulte in the North-east, as well as the A29 highway in the south.

-Micro-region 4: 27 excavations were carried out in this region, most of which were done prior to the construction of the A29 highway in the Centre-north or of the commercial development zone of Haute-Picardie in the North-west. Very recently and still pending publication, the very large construction project of the Canal Seine-Nord-Europe resulted in the partial or full excavation of several dozens of sites, 9 of which are newly discovered Roman rural settlements in this micro-region alone, 5 of which are post-built farms (fig. 9). This last indication might give a better idea of the proportion of post-built sites which remain undiscovered, even for the Roman period. Therefore, table 1 above clearly shows that the current knowledge of rural settlements does not represent accurate proportions of the rural settlements which would have existed.

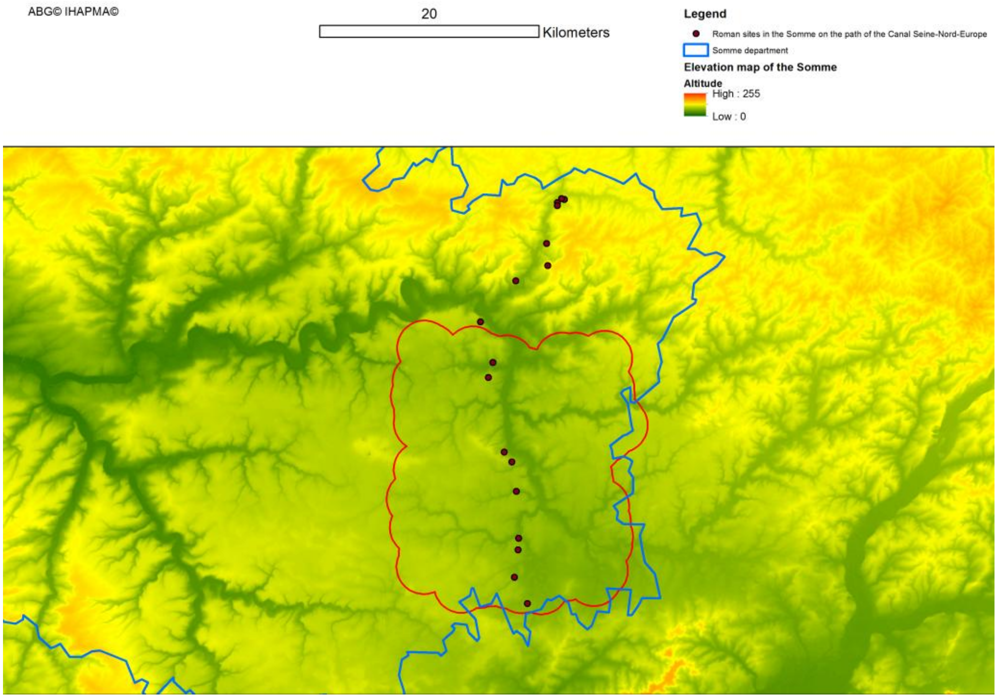


Figure 9: Roman sites in the Somme on the path of the Canal Seine-Nord-Europe, ABG©

1.5.2 Contextual dataset

Regarding urban contexts and secondary foci in the landscape (sanctuaries, secondary agglomerations), these follow very different functional and locational patterns compared to rural settlements. They furthermore should be studied on an individual basis, as they can span several dozens of hectares on diverse and sometimes unique topographical settings. Thus, farms and villas will receive the full focus of this study, but the other categories of sites will still serve a necessary purpose: the location of rural settlements certainly depends on the neighbourhood of urban, commercial or religious centres.

The main cities were built during the Gallo-Roman transition period between 50 BC and the beginning of the 1st century AD. These *capita civitatum* were the administrative, economic and commercial centres of their respective *civitas*, which they controlled and taxed, but they could also be religious and military centres (Collart *et al.* 2004). Only Vermand and Thérouanne have not retained the main role in their local administrative division and are now small villages, while all other cities have maintained an urban character to this day.

Among local foci, sanctuaries are generally large complexes including one or several stone-built *fana*, the typical Gallo-Roman temple, with its concentric square plan of Late Iron Age origin (Agache 1997; Fauduet *et al.* 1993; Fauduet 2010). They may in exceptional cases contain monumental 'classical' temples, theatres and baths, such as in Ribemont-sur-Ancre (Amandry *et al.* 1999). Secondary agglomerations often include sanctuaries amidst their sometimes very large superficies (Petit *et al.* 1994; Pichon 2012; Pichon 2006; Favory 2012; Deru 2012; Petit 2000) and can also be located in unique settings, such as the very large urban and religious agglomeration of *Briga* at Eu, in the South-western margin of the *civitas ambianorum*, which is located on a very prominent and forested area (Mantel *et al.* 2006). Local foci also display very long occupation durations – on average 381 years – as indicated in the graph below, where they are shown to be occupied during durations from 100 to 800 years (fig. 10).

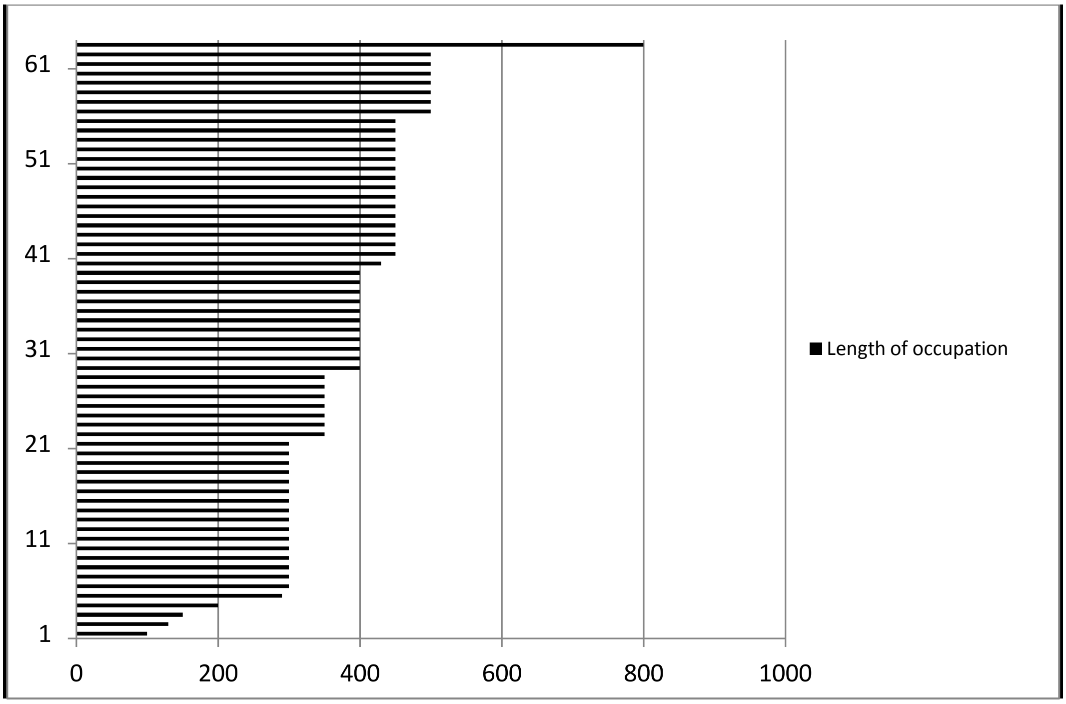


Figure 10: Distribution of maximum occupation duration of secondary agglomerations and sanctuaries, ABG©

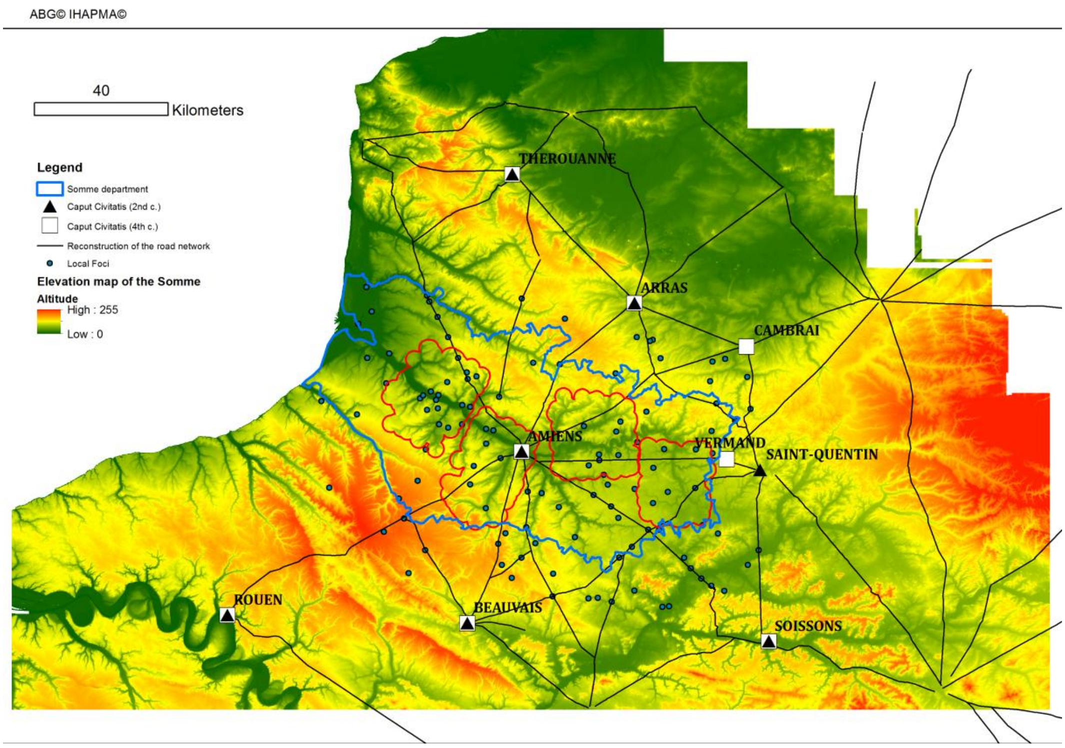


Figure 11: Contextual dataset in and around the studied micro-regions, displaying the Roman roads, the *capita civitatum* in the 2^{nd} and 4^{th} centuries AD, as well as the secondary agglomerations and sanctuaries (local foci), in and around the Somme department, ABG©

This contextual dataset assembled and depicted in figure 11 enables us to evaluate the importance and influence of regional and local focal points towards rural settlement patterns. Burials and sparse finds will not be considered, as they are very punctual elements of the landscape and are either parts of rural or urban settlements, or simply isolated finds which are of no influence and had relatively no dependence towards rural settlements. It should also be noted that no fortifications are included in this dataset, as they were all abandoned shortly after the Augustan consolidation of the region, at the end of the 1^{st} century BC.

1.5.3 Cartographical data

Concerning the primary spatial data, it will be processed in ArcMap 10.4 © (ESRI), which is currently the most frequently employed Geographic Information System (GIS) creation and management software. The geographic data includes a vectorised administrative map of the department (free national data) as well as DEM rasters produced by the IGN (National Geographic Institute) at two pixel resolutions (BD ALTI© 25m and 75m). The DEMs are elevation models interpolated from the RGE ALTI© 1m LIDAR elevation data, as well as the photogrammetric combination of scanned IGN paper maps.

Reconstructions of the palaeolandscape have not been carried out yet in the Somme, therefore one cannot rely on the only land cover maps available: the European Corine LandCover 1990, or the forest cover map produced by the EFDAC (European Forest Data Center). These European scale datasets lack precision considering the modelling approach. Furthermore these contemporary variables are unstable through time and can therefore only be used towards Cultural Resource Management, because site location analysis requires stable variables or reconstructions of historical ones. Therefore these land-use and forest maps will not be used. The vectorised archaeological datasets are the property of the ABG program: the Roman road network (observed and reconstructed), ancient administrative regions, rivers and all point features (all types of sites).

This modelling method will thus mainly stem from the transformation, analysis and combination of the DEM, the sites, the roads and the rivers. Historical soils and land-use – which potentially are the most useful and influential variables for site location – are absent in the department, as are geomorphological reconstructions or even contemporary pedological data. This is a concern shared by many researchers in France, where soil data has not yet been systematically acquired and distributed nationally, as opposed to the Netherlands.

This presentation of the dataset concludes the introductory excursion into the subject under examination and now will be explained the purpose of this thesis by presenting its research questions.

1.6 Research questions

The purpose of this thesis is to use predictive modelling in order to analyse settlement patterns and site location in the Roman Somme. Therefore, the main question which this models will aim to answer is: According to site location analysis using predictive modelling, which environmental or human factors influenced the rural settlement patterns of the Somme during the Roman Imperial period?

This very broad question will be divided into sub-questions, which specific models will try to answer. The first group of questions is concerned with the physical environment. The topography is not very contrasted in the area, but some steep slopes are still present in many valleys. The landforms of this thesis' micro-regions are not uniform at all, displaying many valleys, plateaus, alluvial plains of varying orientations and suitability for settlement. How do topographical characteristics of the landscape influence rural site location? This very important question cannot be solved entirely without accurate reconstructions of the past soils and fluvial parameters – which cannot be carried out here – but a correlation between site location and landform types, slopes, orientations and exposure to the sun might shed light on the relation between Romans and their physical environment.

The hydrographical properties of the landscape are a recurring element of site location analysis. As the name of the department implies, the entirety of this study area is part of the basin of the Somme. This does not mean that all sub-regions have equal hydrological characteristics: many sites clearly depend also on tributary rivers, some of them being relatively substantial and navigable, even more so during the Roman period. Indeed, some sites such as the large villa of Frémontiers display small fluvial docks (Revert 2016, 87; Ben Redjeb 2013, 425–427). While there are some benefits to the proximity of a rural settlement to a river, especially on the gentle slopes and terraces of mid-valley, some disadvantages may outweigh them. Indeed, many areas were and still are prone to important floods, as well as being quite marshy, particularly in the valley floor of the Somme. One might thus ask if valley slopes and terraces were a preferential area of rural settlement, or if drier areas such as the plateaus and its dry valleys were better suited?

These environmental factors – if proving relevant and efficient – will be combined with socio-economic human factors that potentially constituted the most important role in conditioning the settlement patterns. While the proximity of a Roman road is already considered a beneficial factor by most researchers, does it really influence the location of settlements so that they are skewed towards its proximity?

Also cited as important factors in every monographic description of Roman sites and excavations, the influence of the vicinity of cities and regional markets is never actually quantified. Does the proximity of a city, a secondary agglomeration or a religious agglomeration actually benefit the location of rural settlements?

Through the modelling process and the ensuing analysis of site location, answering these questions should help understand which choices or parameters helped create the settlement patterns of the Roman Somme.

1.7 Thesis plan

The general archaeological and geographical context having been introduced, as well as the problems, aims and questions underlining this work, the remainder of this thesis will follow a simple tripartite plan. Chapter 2 will describe the methodological context and approaches underlying this thesis. Chapter 3 will follow by a description of the modelling process and its results, while Chapter 4 will allow a discussion of the relevance of this methodology and the signification of the results towards answering the research questions.

The methodological Chapter 2 will first introduce the principle of predictive modelling and its development, before giving more details about the Dutch context of predictive modelling from which this methodology is strongly inspired and finally focusing on the specifics and limits of the latter.

The modelling chapter 3 will then detail all the processes involved in the modelling of each variable and its combinations with others. The environmental variables and their combinations will be explored first, followed by the results of the different models as well as their efficiency and predictive power. Afterwards, the socio-economic variables will follow the same workflow.

The final part of this thesis' framework in Chapter 4 will be concerned with the interpretations of the models' outcomes, in terms of their added value to site location analysis of the Roman Somme, as well as their contribution to the wider field of archaeology of northern Gaul on the one hand and to archaeological predictive modelling on the other hand.

A final conclusion summing up the issues, positive outcomes and perspectives of this research will close this thesis.

Now will therefore be opened the development of this thesis with the following methodological chapter.

Chapter 2. Methodology

2.1 Principles and first developments of predictive modelling

2.1.1 Definition

The simplest and most commonly cited definition of archaeological predictive modelling is the following: to 'predict the location of archaeological sites or materials in a region, based either on a sample of that region or on fundamental notions concerning human behaviour' (Kohler *et al.* 1986, 400).

Included in this definition are the two main goals of this field of study. The first one involves the prediction of the location of sites based on an already existing sample of sites and is therefore very extensively used in Cultural Resource Management (CRM). The second one models probable site location based on theoretical assumptions on the relation of humans with their landscape (Wheatley 2004, 5). While the second approach only – also known as Archaeological Location Models (ALM) – will be used during this thesis, it nevertheless stems from the first one, whose methods of model-building and evaluation are shared with ALM. Although archaeological predictive modelling has long divided these two approaches in terms of inductive and deductive predictive modelling or data-driven and theory-driven predictive modelling, the recent years' developments have shown that no methodological gap actually exists between these approaches which should benefit from each other and which will now be detailed (Verhagen and Whitley 2012, 90–91).

2.1.2 First developments

Since Processual Archaeology took hold in the late 1960's, the correlation between human settlement behaviours and their physical environment – already acknowledged by previous studies – started to be systematically inquired through quantification methods (Verhagen and Whitley 2012, 51). On this basis, archaeological predictive modelling was born in 1975 in Wyoming and Colorado in the United States, through the CRM-driven projects such as the 'Glenwood Project', led by Kvamme and his colleagues (Morris *et al.* 1975; Morris *et al.* 1979; Burgess *et al.* 1980; Kvamme 2006, 8–9). Their methodology consisted in correlating specific variables of the physical landscape with sample sites, only employing calculators, thereafter transcribing the statistical relations on hand-drawn maps. The latter were then relatively successfully used by the Land Management Units who ordered these studies.

Predictive modelling was first applied to site location analysis in 1976, in the pioneering study of Jochim, which is often considered as the birth of Archaeological Location Modelling (Jochim 1976). His approach was mainly a non-spatial one: complex mathematical equations were formulated to model the ecological parameters on which hunter-gatherers were supposed to depend, revealing their caloric potential and demographic expectations. Survey data fed these variables which were mainly extrapolations of the former (Jochim 1976, 33). Only after all the statistical results were gathered, Jochim translated these in the form of hand-drawn maps. This first methodology was an extremely painstaking one, which was supported by an enormous amount of survey data about the archaeological sites, their physical environment, and the complex ecological systems around which hunter-gatherer societies revolved.

In the 1980's, Cultural Resource Management-oriented predictive modelling took a very significant leap forward, both in terms of number of practitioners and methodological approaches (Kvamme 2006, 10–12; P. Verhagen *et al.* 2010, 430). Multivariate statistical approaches were first developed in Arkansas during the FORTRAN program in 1980-1983, which employed computer programs, but no GIS yet (Kvamme 1983). They could already compute and map variables that are still some of the most employed in current predictive modelling: 'slopes, elevations, aspects, local relief, ridges, drainage lines, terrain variance and distance surfaces for stream vectors' (Kvamme 2006, 10). Still, predictive modelling was oriented towards a 'landscape as now' rather than a 'landscape as then', because its goals were concerned with the contemporary setting only (Harris *et al.* 2006, 43).

GISs though were not yet used, as this first came to be in 1985 through the impulse of the Society for American Archaeology (Kvamme 2006, 11), the results of which were published by Judge, Sebastian and many colleagues (Judge *et al.* 1988). This latter publication, although nearly three decades-old already, constitutes a very complete and rigorous methodology on which any spatial predictive modelling endeavour should strongly take inspiration. It constitutes the first methodological and historiographical synthesis of archaeological predictive modelling and discusses widely the issues and critiques imparted to predictive modelling, through the dichotomy between the so-called inductive and deductive approaches.

From this solid starting foundation, predictive modelling was then spread to Europe and mainly applied towards CRM and development planning, where ‘it has become a multi-million dollar industry’ (Kvamme 2012, 337), similar to the United States. The Western part of the United States constitutes in fact an ideal ‘playground’ for inductive predictive modelling: there are vast uninhabited areas with an excellent preservation and visibility of archaeological structures, thus allowing test surveys to be carried out, which in turn gives the opportunity to verify the predictive power of models, i.e. their ability to predict site location accurately and precisely (Brandt *et al.* 1992, 270). Already in 1986, 70 papers and essays pertained to predictive modelling (Kohler *et al.* 1986; Harris *et al.* 2006, 43).

2.1.3 The inductive/deductive dichotomy and Post-Processual critics

CRM-oriented models tend to be conceived towards a goal of efficiency, not in archaeological terms but financially, and are therefore sometimes statistically and theoretically deficient (Verhagen *et al.* 2009, 6). Recurring issues include the falsity of low-probability site densities, inducing unexpected finds in ulterior surveys and excavations. Completely inductive approaches applied until recently saw the production of a probabilistic model as an end-product and correlation was therefore wrongly taken as a causal-effect (Verhagen and Whitley 2012, 50–51).

Correlation is a very important statistical principle in predictive modelling: it defines a relation between two or more numerical values, in the present case proportions of sites and proportions of expected sites, i.e. the number of sites in a specific area if the distribution of sites was completely uniform. But correlation is not synonymous of influence or causality, as no statistical model could possibly prove these. One telling example is given by Lock and Harris, who recount the modelling of a predictive map in Western Virginia, in which the highest correlation with human settlement was produced by the contemporary distribution of sycamore trees (Harris *et al.* 2006, 49). In this model, although the correlation was very high, there was no deterministic causality to be found between current trees and ancient settlements. Though, both of these variables may be correlated with a combination of variables which allows them to exist in some areas rather than others.

The way in which most scientific disciplines can transform a correlation into a causality or influence, be it deterministic or partial, is through reproduction of these parameters and their expected results. The archaeological discipline being defined by the impermanence and non-reproducibility of its data, such conclusions may never be attained. Despite this, strong archaeological correlations in statistical models, when repeatedly and rigorously put in evidence, considerably strengthen a theoretical hypothesis or claim, giving it a formal, spatial and statistical weight and consistency which could help formulate and consolidate hypotheses on site location and may also give more value to a predictive assessment in the case of CRM.

During the 1990's and the early 2000's, the discipline continued to experiment with many statistical methods of model-building, testing and evaluation. This happened despite the very strong critics imparted by some of the post-Processual archaeologists, which led predictive modelling to a lengthy scholarly defence which is not entirely solved to this day (Kvamme 2006, 11). Among the most prolific and constructive of these critics is David Wheatley, who incidentally has published predictive modelling studies. His scepticism was especially directed towards correlative predictive modelling, i.e. models which are built on the basis of site data on top of being tested by them.

The main criticisms are concerned with the 'ecological fallacy or environmental determinism', in which the behaviour of humans is simplified to obedience towards the physical environmental factors, thus removing any role from the historical and cognitive aspects of ancient life (Wheatley 2004, 6-7). At first, the defence of the practitioners of CRM-oriented predictive modelling indicated that this bias towards environmental determinism is in itself caused by the type of spatial data used in GIS, which, as noted by van Leusen, focuses on 'soils, topography, geology, hydrology, the digital terrain model and its derivatives' (Gaffney *et al.* 1995, 368). CRM-oriented modelling was also justified by van Leusen as 'perfectly valid to try to hunt down environmental correlates of settlement location (...) as long as no simplistic causal "explanation" is attached to these correlations" (Gaffney *et al.* 1995, 369). That statement is partially valid, although it does not entirely validate CRM-modelling, as argued by Wheatley and Gaffney (Wheatley 2004, 7; Gaffney *et al.* 1995, 373). Furthermore, the latter indicated that many CRM-oriented predictive models' performances are poor, and therefore do not 'actually work very well' and should stop being created (Wheatley 2004, 8-10).

The 'deductive' or explanatory approach too was criticised, as only employing deduction in the modelling process is synonymous to translating one's idea into a spatial model, with no correlation or test with the archaeological documentation. Such models were therefore – and justifiably – deemed as lacking predictive power and only able to help understand 'trivial cultural processes' (Kvamme 2006, 11). Predictive modelling in general was also criticised for its 'inability' to deal with complex systems, because of its statistical paucity, its lack of a solid theoretical ground and because of the feeble quality of the archaeological data used in the modelling process. Due to this unsolved scholarly conflict, the inclusion of predictive modelling in archaeological academic research has long been halted (Verhagen *et al.* 2009, 20).

To strengthen the scientific solidity of 'inductive' predictive modelling, a very wide range of statistical methods have been developed and experimented. Density transfer can be cited – which simply extrapolates the density of sites of a sample area to analogous environmental areas – or many other methods which will not be detailed in this chapter, but are listed by Kvamme (Kvamme 2006, 23). The most commonly employed method in archaeological predictive modelling is multiple logistic regression analysis: multiple independent variables (possessing a range of categorical values) are compared with a dependent variable, i.e. possessing two classes, generally the presence or the absence of sites, or simply put the distribution of sites (Warren 1990, 99). The product is the probability of an independent variable to be part of one of the classes of the dependent variable, therefore to display site presence or not, be it on a scale of 0 to 100, 0 to 1 or 1 to 3.

Despite these new developments, the post-processual critics and the slow changes in practices still kept predictive modelling far from the state-of-the-art theoretical frameworks and academic fields of archaeological research, especially in the recent field of landscape archaeology. The latter was developed on the impulse of post-processual authors, whose theoretical ambitions were more related to ‘space’ than ‘place’, with symbolic context rather than site (Tilley 1994; Harris *et al.* 2006, 43–44).

Nonetheless, landscape archaeology, just as predictive modelling and geographical analysis in general, is confronted with internal issues regarding the role of the physical environment and of socio-cultural aspects. Landscape is in itself a very vague term which ‘permeates everything’ (Bender 1999, 5), and ‘whose potential range is all-encompassing’ (Morphy 1993, 200). As put by Pickering, ‘Landscape is a concept in which the intention is known but, as yet, still remains undefined’ (Pickering 2003, 39). In its most extreme post-modernist forms, phenomenological landscape studies can completely disregard scientific inquiry into any kind of evidence, which is the cause for an important counter-phenomenon inside landscape archaeology, at odd with its most ‘mystical’ elements (Fleming 2006, 270). This fierce theoretical debate between proponents of ‘hyper-interpretive’ literary productions on one hand (Fleming 2006, 275–276), and those of scientific analyses on the other hand, had comparatively many more peaceful ripples in predictive modelling, where practices and pragmatic choices generally condition one’s method of inquiry.

Predictive modelling thus had to find a way to attenuate this omnipresent theoretical dichotomy and dualistic scholarly opposition of cognitive aspirations and functional practices, this middle ground being the focus of the next paragraphs.

2.2 Solving the dichotomy

2.2.1 Finding a middle ground

As put by Kvamme: ‘The archaeological dichotomy that has arisen claiming distinct correlative and deductive approaches to modelling is an unfortunate historical accident; they need not be different but can and should be one and the same’ (Kvamme 2006, 13). Whitley simply brushed aside this entire nomenclature and replaced it by ‘correlative’ and ‘cognitive’ approaches, which better define the process in which the models are built: correlative models are those which use site data to be built (commonly accepted as ‘inductive’), whereas cognitive ones are built on the basis of hypotheses regarding site location and are then tested with the help of site data (Whitley 2004a, 5; Whitley 2005).

Combining so-called deductive and inductive approaches may therefore help reconcile predictive modelling with theoretical discussions on landscape archaeology. Since the end of the 1990’s, ‘deductive’ approaches received some efforts which were focused on trying to give it an expert judgement building process, while also allowing a statistical testing of the results on the basis of archaeological data. In regards to site location analysis, Whitley formulated a theoretical foundation on which the present methodology will be strongly based. He describes ‘archaeological probabilistic modelling’ as based on three assumptions: people make decisions on where to settle, these decisions are correlated with the physical, socioeconomic and cultural environment, and finally the models of the latter can be confronted with site data and explain settlement patterns (Whitley 2001, 2; Whitley 2004a, 4–6). On this theoretical basis, one must then design how and why decisions are taken, which variables affect them and how the influence of these variables can be measured (Kohler *et al.* 1986, 432–440; Whitley 2001, 6).

The inclusion of new statistical methods to the field is a very important aspect of the new research-oriented deductive modelling, as it can now deal much better with bias and efficiency (Verhagen and Whitley 2012, 55).

For this endeavour, it is mandatory to give explicit and transparent quantitative weights and relations between the variables involved in the modelling process. This semi-quantitative approach is called ‘multi-variate/criteria analysis’, which consists in the algorithmic combination of variables weighted on the basis of expert judgement. Then the correlation between these variables and the actual archaeological data can be made using varied statistical methods. This weighted multivariate approach is in fact a middle ground between the inductive/deductive dichotomy. The weighting equation, which is a formal translation of one’s theory regarding a research question, gives a much needed transparency and replicability to such models, just as stronger statistical methods can improve and evaluate the robustness of the model and quantify its uncertainties (Verhagen 2012, 311).

As such, this new array of techniques is only beginning to be employed, refined and tested. This lengthy gap between theoretical discussions on cognitive archaeology and their actual application in predictive modelling projects was already noted in 1992 by Wheatley, who concluded on a general lack of contact between theoretical archaeologists and the practitioners of GIS (Wheatley 1993, 137).

2.2.2 The socio-cultural approach

The diversity of approaches in predictive modelling, even when only considering ALM-oriented research, can lead to a very critical – sometimes sceptical – view of the practice. As was noted by Lock and Curry, the divide between the study of space and the study of place is not yet closed, which is exemplified by the lag between the many theoretical studies trying to bridge the gap and the much fewer applications of such ideas (Curry 1998, 143; Lock 2001, 160).

To improve the theoretical framework in predictive modelling and to get back on track with cognitive inquiries into human behaviour, it is fundamental that socio-economic variables be included into the modelling process. Although some attempts were published in the late 1990’s and early 2000’s, using visibility of the landscape as a deterministic precondition to the location of fortified sites in the Croatian island of Brač, only very recently have many more aspects been applied to predictive modelling and tested successfully (Stancic *et al.* 1999; Verhagen *et al.* 2011). Site location analysis, just as any analysis of human remains, should acknowledge the complexity of human decisions – or the lack of any pattern to uncover them in some cases – but still provide well-supported explanations for the observed patterns in data, which should relate as closely as possible to the historical behaviours, as defended by Hodder (Hodder 1992, 14–16).

The early works carried out in Croatia have been expanded considerably through the Dutch and French works of Verhagen, Nuninger, Bertoncello and other colleagues, who have especially contributed to numerous attempts at modelling the relation between site location, their natural environment as well as the human environment (Verhagen *et al.* 2007). This allowed a bridge to be built with an important project in French spatial archaeology: ArchaeDyn (Spatial Dynamical Analysis from Prehistory to the Middle-Ages) which originated from Archaeomedes, a previous long-lasting European project concerned with ancient settlement patterns (Favory, Nuninger, *et al.* 2008; Favory and Nuninger 2008b; Favory and Nuninger 2008a; Nuninger *et al.* 2007; Nuninger *et al.* 2012; Bertoncello *et al.* 2008; Nuninger 2002).

Led by P. Verhagen, two successive post-doctoral projects have aimed at including socio-cultural and socio-environmental factors in predictive modelling and modelling spatial and economic dynamics. The first one, the IHAPMA project (Introducing the Human (f)Actor in Predictive Modelling for Archaeology), was carried out from 2009 to 2012 (Verhagen *et al.* 2013, Nuninger *et al.* 2015, . The purpose of this project was to implement in three study areas (Argens-Maure and Vaunage in Southern France and Zuid-Limburg in South-Eastern Netherlands) a ‘protocol that can be easily implemented for different regions and time periods, using contextual analysis, statistical comparison and predictive modelling of site location as the primary tools to gain a better understanding of cross-regional diachronic patterns of occupation’ (Verhagen *et al.* 2015, 623). The second project is *Finding the Limits of the Limes* and builds upon the latter methodological work towards a more integrated study of spatial dynamical relations in the Roman landscape of the Lower Rhine. As such it explores new ways to assert the chronological aspects of sites (Verhagen, Vossen, *et al.* 2016), the economic network between rural settlements and military ones (Groenhuijzen *et al.* 2015), or demographic theories on the local populations, both of humans and domesticated animals (Verhagen, Joyce, *et al.* 2016; Joyce *et al.* 2016).

2.2.3 The present state of predictive modelling

In what Kvamme nicknamed in 2006 the 'Second Age of Modelling', the present state of predictive modelling is characterised by a willingness to widen the field of ALM-oriented predictive modelling and construct better controlled variables which incorporate more social elements (Kvamme 2006, 21). This marks the recent publication of a series of papers dedicated to the implementation and testing of specific variables, concerning case studies in France and the Netherlands. Researchers of these two countries have in fact produced the vast majority of study cases of ALM modelling, and are also at the forefront of actual implementations of 'cognitive' and socio-economic variables involved in theory-driven modelling projects, following the earlier theoretical advances called for by British, Dutch and American scholars such as Kvamme, Kammermans, Wheatley, Gaffney and Whitley (Kvamme 2006; Whitley 2005; Wheatley 2004; Gaffney *et al.* 1995; Kamermans 2008; Verhagen *et al.* 2007). The first aspects to have been incorporated were the accessibility and the visibility, through the already well developed least-cost path densities and viewshed analysis (Verhagen *et al.* 2013). These two aspects aim to reconstruct the potential benefits of movement from and towards potential sites, therefore improving their site location probability.

Land-use heritage was very recently experimented with, targeting the definition of the land-use weight of each area per century, i.e. the probability of a site creation based on the density and duration of occupation of previous settlements in a specific area, be they still extant or not (Nuninger *et al.* 2016; Verhagen *et al.* 2015). This sub-model merges modelling theories about the neighbourhood of sites as well as their chronological development, allowing the modelling of space around sites rather than the sites alone. In this aspect, a very strong bridge with the targets of landscape archaeology is beginning to be built. Not only is space part of the modelling process, but also its dynamic throughout time, departing from the fixed and stable conceptualisations of the environment as were designed in past CRM-oriented models.

As well as improvement of the theoretical background and objectives of archaeological predictive modelling, its most technical aspects are also at a turning point. Indeed, the democratisation of relatively user-friendly GIS softwares – in particular ESRI's ArcGIS© or opensource freewares such as QGIS – has led spatial analysis, spatial modelling and digital cartography in general to take a prominent and sometimes systematic place in archaeological practice as well as research, but the same can be said in every space-related discipline. With the exponential increase of computational power, more and more data can and should be employed, subdivided, reproduced and analysed. The resulting complexity of data is the focus of a staggering diversity of international and national interdisciplinary conferences related to digital data in general.

Regarding the use of GIS in predictive modelling, just as in image-based analysis the focus is shifting towards automatisation. Creating tools and protocols to ease the modelling process and especially the statistical elements of the workflow has become imperative. To give an example, to study the correlation of a single model with different sample regions in different periods requires first the creation of as many dependent variables (point features of sites) as there are periods multiplied by micro-regions. Each dependent variable will then have to be evaluated individually against the independent variables and if necessary, with each one of them separately. Even though computational capacities have increased tremendously, the time-cost of the modelling process is still great.

To help solve these problems, initiatives are being developed, notably in the Netherlands, more precisely through the impulse of Verhagen, Nuninger, Bertincello as well as Nüsslein. The latter very recently applied a wide set of predictive models to four micro-regions in Roman Eastern Gaul, which constitute a very important inspiration for this work (Nüsslein 2016b, 303–370). By combining spatial dynamical analysis methods developed by the Archaeomedes, Archaedyn as well as RurLand projects (Rural Landscape in north-eastern Gaul, EPHE) and merging them with the socio-cultural and environmental predictive modelling of the IHAPMA project, Nüsslein's doctoral thesis showcases the high applicability and versatility of the modelling protocols developed collectively.

The creation and continuous improvement of an ArcGIS toolbox and its protocols is part of the initiative of this project. Application, adjustment, improvement and feedback on these protocols and tools are also part of an internship I carried-out under the supervision of P. Verhagen, as part of his VINI-granted post-doctoral project on spatial dynamical modelling. Although promising, and an improvement to the modelling practice, this toolbox will not be applied entirely in the present methodology. Indeed, correlation between dependent and independent variables plays a large part in the model-building of these protocols, which certainly improves accuracy of the model, but also skews its results towards a more correlative result. In order to assert the influence of each category of each variable, the protocol also advocates the use of Principal Component Analysis (PCA) and Maximum Likelihood Classification (MLC). These two methods combined seek to dissect potential categorical intercorrelations and to create new categories, independent of any variable. This considerably eases the analysis of each category's correlation with dependent variables, but also has the effect of compressing together values of different variables which in the end can contradict each other inside the same category. A final issue with this approach is the use of focal statistics: in order to model space rather than sites, the mean of a variable's value in a specified radius gives its value to the cell around which the radius revolves. While it certainly renders each cell's value more representative of the radius' value, it also erases any local specificity of this cell. As a consequence, the model tends to be 'smoothened' by this process, both visually and statistically, with much tighter ranges of values.

As employed by Nüsslein (2016b, 303-370), this method does not include formal theories in the model-building process, which means that research hypotheses are tested at the end of the workflow with statistical results, but are not translated into the model, which follows the PCA and MLC's reclassification rather than a manual one. The experimental methodology which will be described below seeks to use some of the benefits of recent developments, but in a way which relates more to weighted multivariate analysis. A personal toolbox was adapted from the IHAPMA one.

2.3 Methods and protocols of this thesis

This methodology follows the second main purpose of archaeological predictive modelling as described by Whitley: ‘understanding past land use or what can be called site selection processes’ (Whitley 2004a, 2). It is also cognitive, explanatory and probabilistic in the sense that hypotheses on human decision patterns are the foundations on which the different models are built and their outcome is non-deterministic (Whitley 2004a, 6–9; Whitley 2004c, 1–2; Gaffney *et al.* 1995, 370–371). The different phases of this methodology will be described below.

2.3.1 Conceptual phase

The first step in any modelling process is a conceptual phase, which contains research questions, the methodology to be employed, the hypothesis which translates specific expectations and the formal theory which transcribes it (Romanowska 2015, 172). The next step is the technical or implementation phase, which in the present case means the actual model construction and statistical testing process, while the last phase, differently from Romanowska’s workflow, concerns the analysis and interpretation of results. The workflow proposed below summarises the entire methodology, which will be explained step by step (fig. 12).

During the conceptual phase, the specific research question attributed to the modelling process is given. A hypothesis is then projected from the research question as an expected conclusion to the model. From this hypothesis is written a formal theory, given in natural language as an explanation of the hypothetical outcome, through the correlation of different independent and/or dependent variables. Then the analytical method is designed in order to model the formal theory.

To the question ‘Are rural settlements preferably located in areas which are topographically amenable?’ it is hypothesised that areas where the landscape displays few topographical obstacles should be influential towards a higher site location probability. The formal theory would then state that the combination of variables (elevation, landform, slopes), reclassified according to the hypothesis, would produce a positive correlation with site location in classes which model positive factors and oppositely for classes of negative factors. This theory would also be transcribed as a weighted equation, in which each variable receives a specific weight. The analytical method would then describe how variables will be modelled.

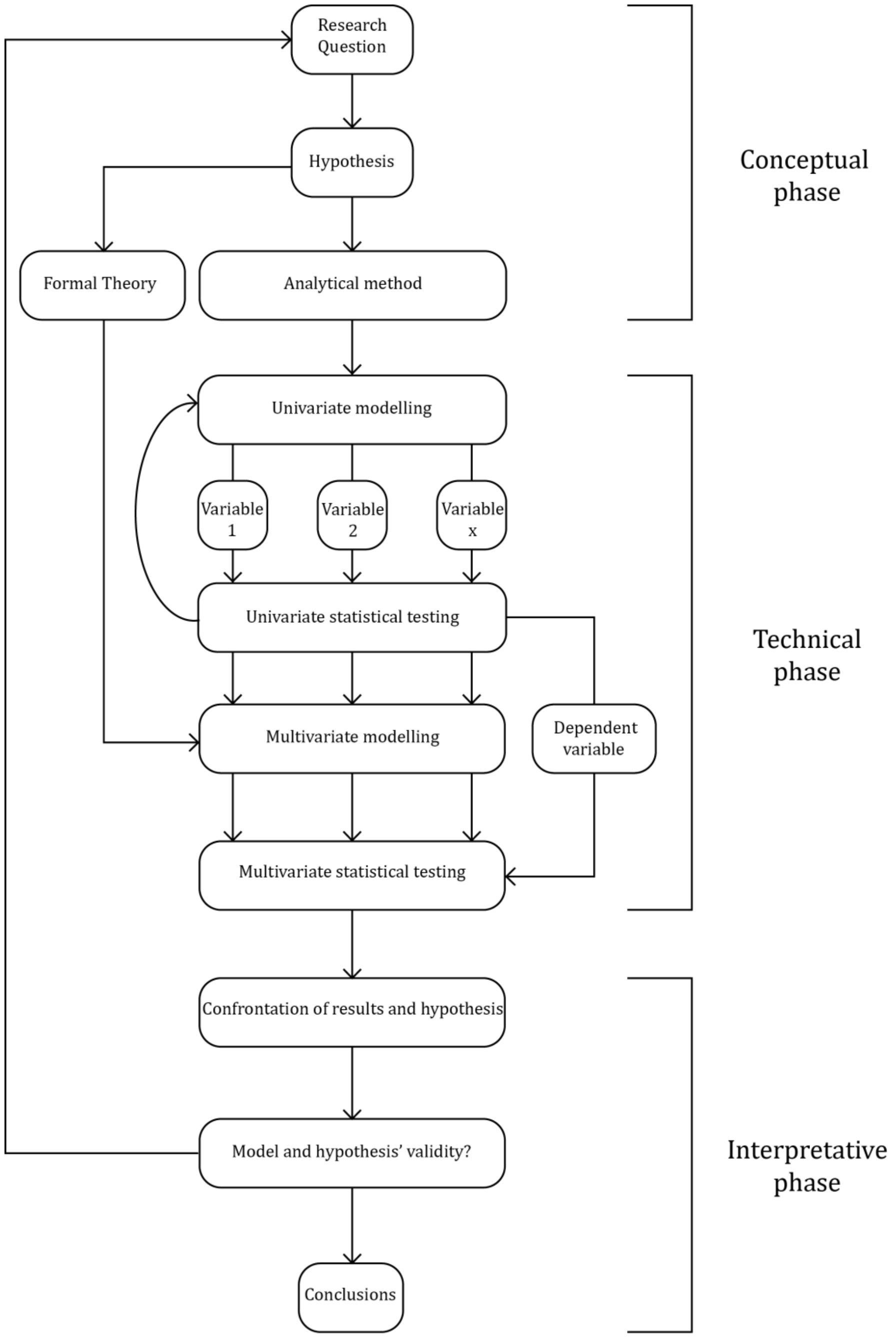


Figure 12: Proposed workflow for this modelling methodology, adapted from the conceptual workflow of Romanowska (2015, 172)

2.3.2 Technical phase

The technical phase is the one which sees the model being implemented through GIS. It starts with the construction of variables, be they dependent or independent. Each independent variable's reclassification is explained, according to the predictive expectation given to each class. These weights are subjectively from 1 to 3 in rank of expected probabilistic correlation with the dependent variable.

Then this univariate model is evaluated with the help of the archaeological dataset. This first consists in stating the randomness probability of the whole model through a Chi-square test (χ^2) This test evaluates the probability of two lists of values to be uncorrelated, therefore if the model shows a random pattern. The two lists of values are the proportions of actual sites per class and their expected proportion in the case of a uniform distribution. If the resulting percentage is above 5% (value of 0.05 on 1), then the model is too random in terms of site location. The lower the values, the higher the probability of non-uniformity and therefore a strong correlation with site location.

The predictive values relate to the relative areas occupied by each class of the univariate model, whose proportion of the total area of each micro-region represents the proportion of sites which would be located in it if their distribution was not influenced by any factor, i.e. in a state of homogeneous site distribution in all areas, which can be due to the contradiction between positive and negative factors, or the lack of correlation with any factor. The actual frequency of sites in each class is then confronted to this homogeneous expected ratio. This is done through three simple calculations which are commonly used in predictive modelling (Verhagen 2007, 48–49): Kvamme's gain, the Indicative value and the Relative gain (table 2).

Kvamme's gain generally produces values below 0 and above 0, with positive values being positively predictive or inversely for negative values. If the indicative value of a class is higher than 1, this ratio indicates a higher site location correlation than in a uniform distribution of sites and the opposite applies if the ratio is lower than 1. The wider the difference between low and high indicative values, the more accurate the model is. The relative gain also produces values below and above 0, although their deviation should not be as large as Kvamme's gain. It allows the assertion of a class’ precision in its prediction, for if this value is important (above 0.20 or below-0.20), it indicates a high precision of the correlation. This statistical evaluation is repeated for each micro-region and each sub-period, therefore with a specific dependent variable each time.

Site and area proportions	**Frequency**	P_o	**Area**	P_a	**Expected frequency**
Signification	Number of actual sites in the class	Proportion of this class’s sites compared to the total number of sites	Surface of the class in km^2	Proportion of the area of the class compared to the total area of the model, equal to the expected proportion of sites in the area	Expected number of sites in each class
Formal equation		Frequency / sum (frequencies)		Area / sum (Areas)	P_a x sum (frequencies)
Class 1	25	0.125 (12.5%)	200	0.50 (50%)	100
Class 2	75	0.375 (37.5%)	150	0.375 (37.5%)	75
Class 3	100	0.50 (50%)	50	0.125 (12.5%)	25
Statistical evaluation of the model	**Kvamme’s gain**	**Indicative value**	**Relative gain**		**Chi-square test (χ^{2})**
Signification	Deviation between expectations and actual frequencies: predictive tendency	Deviation between actual frequencies and expectations: predictive accuracy	Difference between actual frequencies and expectations: predictive precision		Homogeneity between actual frequencies and expected frequencies
Formal equation	1- (P_a / P_o)	P_o / P_a	P_o - P_a		Sum of all classe’s (((Each frequency - Each expected frequency)^2)/Each expected frequency)
Class 1	-3.00 : negatively predictive	0.25 : very accurate for a negative class	-0.375 : quite precise		0.00000 : extremely heterogeneous
Class 2	0.00 : homogeneous	1 : homogeneous	0.00 : homogeneous		
Class 3	0.75 : very predictive (maximum of 1)	4 : very accurate	0.375 : quite precise		

Table 2: Site and area proportions, expected values and methods of evaluation of the models’ accuracy, precision and efficiency, IHAPMA©

If the statistical evaluation of the univariate models indicates problems of lack of representation of classes, the model building process may start again. Otherwise, the multivariate weighted analysis can begin: the univariate models are combined together in the way ordained by the formal theory’s formula. The same statistical evaluation as described earlier is then applied again.

2.3.3 Interpretative phase

Comparing the original hypothesis, the spatial and statistical results, different levels of accordance or discordance can be reached. The results are very often contradictory to the starting hypothesis, but if the model is still statistically relevant and even if it is not, this sheds light on the correlation of variables in some areas or others, during one period or another or the lack of any correlation, which may also be a valid conclusion for the model.

From this point onwards, the relevance of the model is described and criticised, before reaching a general conclusion answering the original question. A new hypothesis may then be formulated, taking note of the conclusions of the last model.

For all phases of this workflow, the toolbox and protocols from the IHAPMA example – specifically adapted for this thesis project – allow the automatisation of a few processes, but generally a more controlled interface in which one can progress towards the technical phase: raster and point extractions, division of feature-sets, reclassification of variables, statistical testing, multivariate combination and other steps which are specific to each variable’s construction.

Now the results of this thesis will be presented.

Chapter 3. Models

3.1 Presentation of the results

Of the following models, first the research question will be presented, then the hypothesis and the theory which constitutes the foundation for the model. The model's variables will be justified and explained in terms of classification, weights, and individual relevance. The results of a single-variable modelling test will be produced as well as tables and receive a primary interpretation. That way, it will be easier to assert the influence of each factor in the multivariate weighted model.

Afterwards the equation responsible for the production of the multivariate model will be proposed. For the latter, as well as the individual variables, a map will be produced alongside the tables containing the results of the statistical evaluation of each class. The presentation of the tables is slightly different from the work of Nüsslein (2016b, 317-354): the classes are individually presented, rather than the model as a whole, but the elements included are similar. Each table contains the details concerning each micro-region and they show three categories of elements. Presented first is the information regarding the predictive classes and their actual content, as well as their weight and the area ratio they represent, which also constitutes the ratio of expected sites (P_a). The second category of the tables includes the number of observed sites, the ratio it represents among the total number of sites (P_o) and the number of expected sites according to the total number of observed sites. The last elements concern the predictive values, which assert the gain, accuracy and precision of the model: Kvamme's gain, the indicative value for each class, their relative gain and the global Chi-square test. For the socio-economic variables, which will be combined with the physical environmental model, the tables will include for each predictive value the statistical change brought by the new variable.

Thereafter, every model as a whole will be criticised, with the conclusion forming the starting point for the creation of the next model, which can be an improvement over the precedent or a radically new one. Whenever a model deemed highly relevant is produced, the results will be expanded to assert the evolution of the model's influence in different periods.

First, several sub-models will be concerned with the influence of DEM and hydrology-derived variables: slopes, aspect, solar radiation, landform and distance to rivers. The most relevant of them will be combined in a multivariate weighted model. Then, socio-economic variables will be added to the most relevant variables among the precedents, tested as bivariate models and finally combined with the precedent multivariate model into a new one. They will include relative or absolute distances to important economical and political focal points: cities, secondary agglomerations and Roman roads. Finally a subchapter will be devoted to the evaluation of the most pertinent models in regard to the three types of rural settlements.

3.2 The physical landscape

For this first model, here is the research question from which the model will proceed: Is site location influenced by topography? The hypothesis is that specific topographical classes play an important role in site location, with slopes playing a second role, distance to water a third place while aspect or solar radiation would still have a very slight influence, despite the relatively gentle topography of the study area. In theory, the best regions for rural settlements would be located in valleys' gentle mid-slopes (below 6%) or plateaus , in areas facing South or South-west. The variables which constitute this topographical model are all derived from the 25m resolution IGN© DEM and the vectorised hydrographical network (ABG©).

3.2.1 Slope

The slope raster is produced through ESRI's ArcGis Spatial Analyst© toolbox, computing the average maximum rate of change in elevation of a group of 9 neighbouring cells. A cell whose 8 neighbours possess the same elevation value therefore has a null slope, whereas any change in a neighbour's value would increase the slope. Choosing a percent-rise value rather than a degree for the slope is an arbitrary choice, easing the further reclassification of this variable. This raster was thus simplified into three classes: from 0 to 2% rise, from 2% to 6% rise and from 6% rise to 82.22% rise.

The first class models flat areas which are easily built and farmed upon, giving no trouble for human or animal movement, while the last class aims to model steep areas impeding movement, construction and animal traction for ploughing, including a hypothetical 9% maximum limit for Roman carts, according to the combination of Raepsaet's algorithm of traction friction (Raepsaet 2002) and Roth's ideal 500kg-loaded two-wheeled Roman cart drawn by two mules (Roth 1999, 212). This 9% limit is sometimes used in least cost path models (Verhagen *et al.* 2014, 80), but is often contradicted by actual Roman road remains which climb slopes up to 20%, although generally keep to slopes below 8% (Herzog 2010, 375; Grewe 2004, 30). Considering the difficulties of ploughing through heavy soils without any pavement whatsoever, or general transport of materials and products on private paths, a critical limit of 6% does not seem too low for the rural settlements. The intermediate range of slopes below 6% and above 2% should therefore have a neutral or slightly negative impact on site location.

The results of the single-variable modelling in figure 13 and table 3 below show important variations from region to region: while the first three regions have some representation of moderate to high slopes (lower weight), the fourth micro-region is nearly entirely constituted of very light slopes, as it is the flattest micro-region of the study area. As such, this fourth region's sites do not correlate as well as the others with the slope classes: its Chi-square test indicates relatively high distribution randomness, just as the predictive values show an inversion of predictive power between the medium and high probability classes. Therefore, any further model including slope factors is expected to behave better in the first three micro-regions. In any case, the distribution of classes explicitly shows the gentleness of the relief, which markedly increases the ratio of the high probability area, but all classes of slopes are still settled, which means that site location is not limited much by the relative slope in any area.

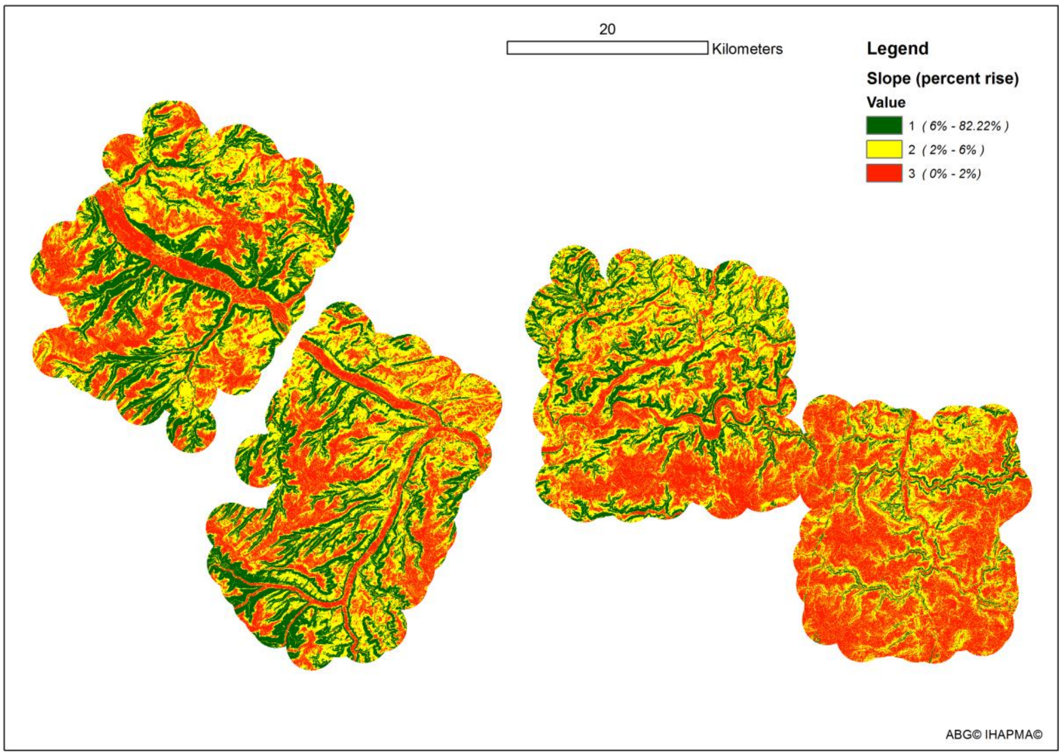


Figure 13: Model of the reclassified slope variable (in percent rise), IHAPMA© ABG©

Slope class	Weight	Observed sites	Expected sites	Area ratio P_a	P_o	Kvamme's gain P_a/P_o	Indicative value P_o/P_a	Relative gain P_o-P_a	Chi-square test
Region 1									
6% - 82%	1	15	30.58	0.15	0.07	-1.04	0.49	-0.07	0.00033
2% - 6%	2	18	28.15	0.13	0.09	-0.56	0.64	-0.05	
0% - 2%	3	175	149.28	0.72	0.84	0.15	1.17	0.12	
Region 2									
6% - 82%	1	10	25.52	0.15	0.058	-1.55	0.39	-0.09	0.00029
2% - 6%	2	16	23.56	0.14	0.094	-0.47	0.68	-0.04	
0% - 2%	3	145	121.92	0.71	0.85	0.16	1.19	0.13	
Region 3									
6% - 82%	1	10	21.83	0.11	0.05	-1.18	0.46	-0.06	0.00249
2% - 6%	2	18	26.74	0.13	0.09	-0.49	0.67	-0.04	
0% - 2%	3	176	155.43	0.76	0.86	0.12	1.13	0.10	
Region 4									
6% - 82%	1	5	6.30	0.03	0.02	-0.26	0.79	-0.01	0.23481
2% - 6%	2	7	12.50	0.05	0.03	-0.79	0.56	-0.02	
0% - 2%	3	227	220.20	0.92	0.95	0.03	1.03	0.03	

Table 3: Results of the univariate model of the slope variable (percent rise) ABG© IHAPMA©

3.2.2 Aspect

Aspect is computed in an analogous way to slope, also through Spatial Analyst©. It is similarly the correlation of a cell with its height neighbours, which produces for each cell a value between -1 and 360, corresponding to a clockwise direction in degrees, where values from -1 to 0 represent flatness. A pre-classification is therefore already implemented in the software, giving a set of 8 directions and a flatness value. Nonetheless, further reclassification is necessary and directions were therefore grouped according to their relative exposure to the sun. In the high probability class those aspects which are the longest and most intensely exposed to the sun during the day (South and Southwest), as they are beneficial for cereal farming in this relatively wet area of oceanic climate. The medium probability area includes the West, the East, the Southeast and flat areas, while the low probability area contains the North, the

Northwest and the Northeast. As was noted earlier, valley-sides facing north are also more prone to steeper slopes, due to differential quaternary ice melting and erosion.

The results of the univariate model (fig. 14 and table 4) clearly indicate that this variable has no correlation with site location. Indeed, all micro-regions show a very high Chi-square result, indicative of significant randomness in the distribution of expected and actual sites, i.e. a near uniform distribution. The predictive assessment results all concur with this conclusion, with incoherent predictive strengths or nearly no change. Another test – on the non-reclassified aspect raster (not shown here) – asserted whether this reclassification was the cause of the univariate model's randomness, but the same behaviour was in fact obtained. Therefore, aspect should be considered as irrelevant for site location in the Somme. It may nonetheless be verified one last time if sun exposition plays any part in site location, through solar radiation analysis.

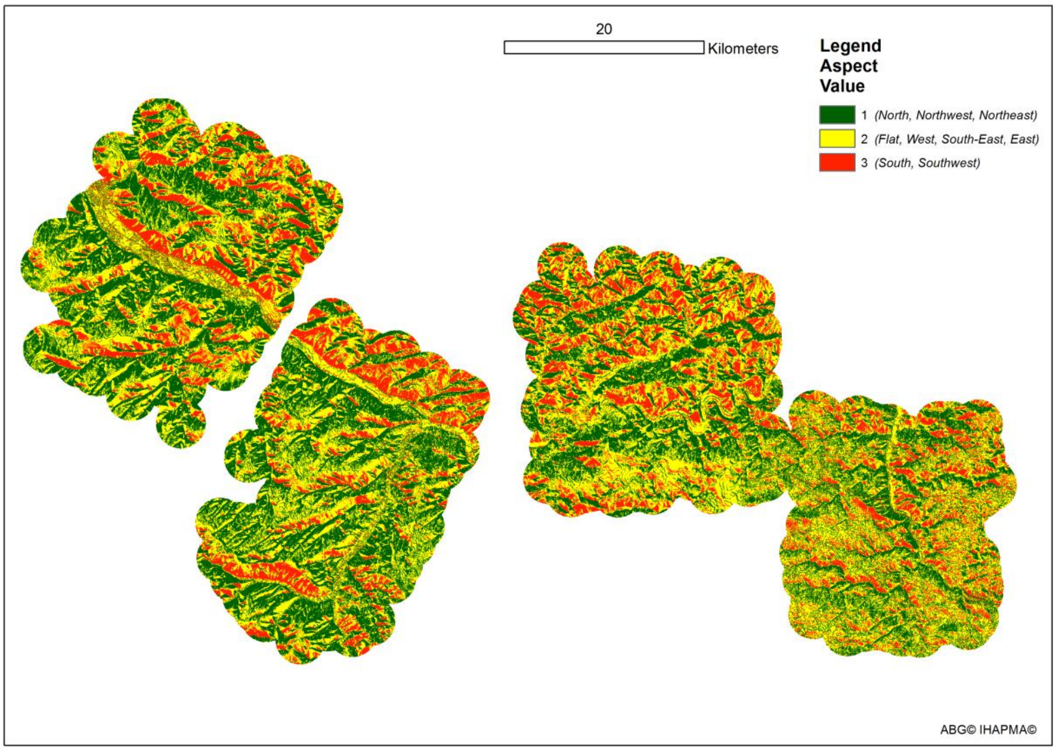


Figure 14: Model of the reclassified aspect variable (in percent rise), IHAPMA© ABG©

Aspect class	**Weight**	**Observed sites**	**Expected sites**	**Area ratio P_a**	P_o	**Kvamme's gain P_a/P_o**	**Indicative value P_o/P_a**	**Relative gain P_o-P_a**	**Chi-square test**
Region 1									
N, NW, NE	1	92	88.05	0.42	0.44	0.04	1.04	0.02	0.79428
F, W, E, SW	2	80	80.64	0.39	0.38	-0.01	0.99	0.00	
S, SW	3	36	39.31	0.19	0.17	-0.09	0.92	-0.02	
Region 2									
N, NW, NE	1	79	75.81	0.44	0.46	0.04	1.04	0.02	0.22511
F, W, E, SW	2	71	65.73	0.38	0.42	0.07	1.08	0.03	
S, SW	3	21	29.45	0.17	0.12	-0.40	0.71	-0.05	
Region 3									
N, NW, NE	1	60	73.53	0.36	0.29	-0.23	0.82	-0.07	0.01416
F, W, E, SW	2	104	83.55	0.41	0.51	0.20	1.24	0.10	
S, SW	3	40	46.92	0.23	0.20	-0.17	0.85	-0.03	
Region 4									
N, NW, NE	1	91	85.45	0.36	0.38	0.06	1.06	0.02	0.72839
F, W, E, SW	2	101	106.38	0.45	0.42	-0.05	0.95	-0.02	
S, SW	3	47	47.17	0.20	0.20	0.00	1.00	0.00	

Table 4: Results of the univariate model of the aspect variable ABG© IHAPMA©

3.2.3 Solar radiation

Through the use of the Area Solar Radiation tool of the Spatial Analyst©, a DEM input can be processed into a radiation raster output. Its unit is the watt hours per square meter (WH/m^2) and it can be computed for different durations, seasons or latitudes. Following the IHAPMA toolbox, the total radiation was computed for only one day, the 80th of the year, from 0h to 24h in the local latitude. Because of the enormous computational weight of this analysis, it was decided to use the 75m resolution IGN© DEM rather than the 25m resolution one, which would have required more than a day to process. The result of the 75m resolution was still relatively satisfactory in terms of precision. The results of focal statistics on the solar radiation raster also ended up with very similar statistical results, so only the model based on single cells is given here. It was reclassified into three quantile categories. Therefore, each category possesses a similar number of occurrences, which should make any correlation with site location more representative.

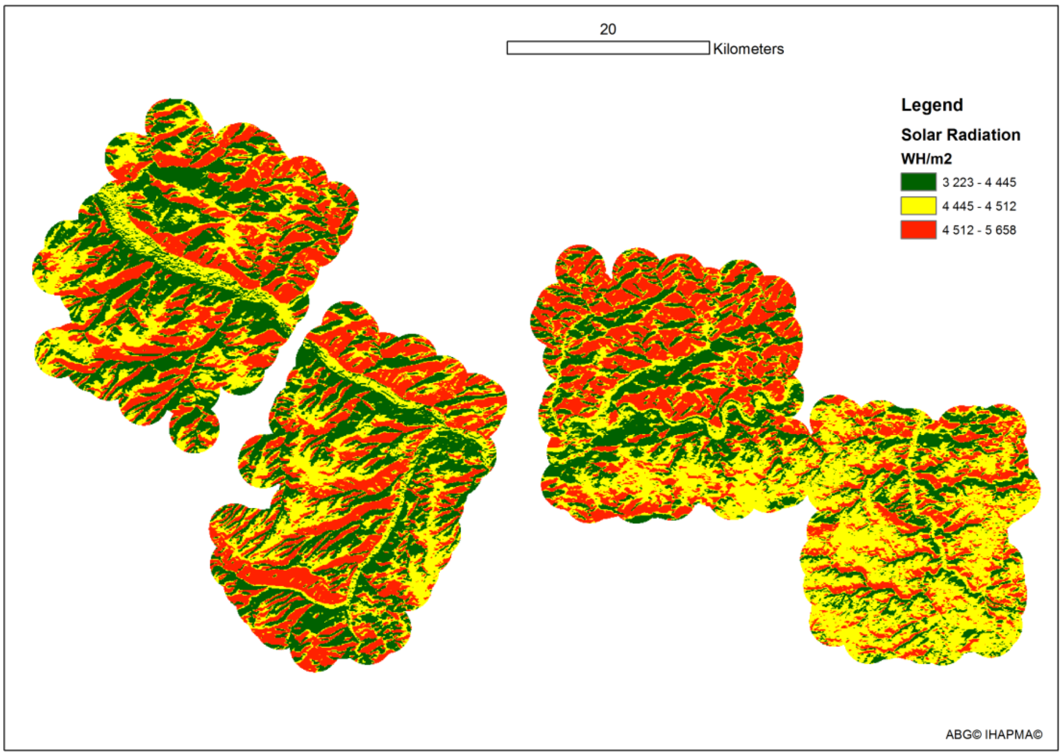


Figure 15: Univariate model of the reclassified solar radiation variable, IHAPMA© ABG©

The results of the univariate model show very contrasting and contradictory results (fig.15 and table 5). Indeed, only region 2 and region 3 seem to show a correlation with solar radiation according to the Chi-square test, but their indicative values do not show any intelligible pattern, with an inversion of the middle and high probability classes which fails to be explained. The first and fourth regions show almost no influence at all from the solar radiation classes, with a near-homogeneity that puts this variable's pertinence in doubt, which is why – just as the aspect variable – it will not be combined with other factors. Although opposite to the precedent expectations, the few classes that show correlation with site location could very well do so because of a common cause, as solar radiation depends on the topographical location of a cell. This common cause may be researched through the next variable.

Solar radiation class	**Weight**	**Observed sites**	**Expected sites**	P_o	**Area ratio** P_a	**Kvamme's gain** P_a/P_o	**Indicative value** P_o/P_a	**Relative gain** P_o-P_a	**Chi-square test**
Region 1									
3223 - 4445	1	80	78.50	0.38	0.38	0.02	1.02	0.01	0.29586
4445 - 4512	2	69	60.93	0.33	0.29	0.12	1.13	0.04	
4512 - 5658	3	59	68.58	0.28	0.33	-0.16	0.86	-0.05	
Region 2									
3223 - 4445	1	57	61.73	0.33	0.36	-0.08	0.92	-0.03	0.00038
4445 - 4512	2	71	48.67	0.42	0.28	0.31	1.46	0.13	
4512 - 5658	3	43	60.60	0.25	0.35	-0.41	0.71	-0.10	
Region 3									
3223 - 4445	1	46	62.86	0.23	0.31	-0.37	0.73	-0.08	0.00038
4445 - 4512	2	88	62.59	0.43	0.31	0.29	1.41	0.12	
4512 - 5658	3	70	78.55	0.34	0.39	-0.12	0.89	-0.04	
Region 4									
3223 - 4445	1	52	50.29	0.22	0.21	0.03	1.03	0.01	0.88160
4445 - 4512	2	133	131.53	0.56	0.55	0.01	1.01	0.01	
4512 - 5658	3	54	57.18	0.23	0.24	-0.06	0.94	-0.01	

Table 5: Results of the univariate model of the solar radiation variable, ABG© IHAPMA©

3.2.4 Landforms

Landforms are considered as one of the most stable parameters of the landscape, related to but distinct from slopes (Kvamme 2006, 20). Reclassifying a DEM would not be a correct approach to model landforms, as the different micro-regions have very different thresholds for each category. To model an approximation of landforms, two main methods have been applied in geographic information systems: Topographic Position Index (TPI) and Derivation from Mean Elevation (DEV), the latter being derived from the former (De Reu *et al.* 2011; De Reu *et al.* 2013; Weiss 2001). TPI is calculated through difference from mean elevation (DIFF), which consists in applying focal statistics around a cell and to subtract the elevation value on this cell by the mean value of cells around it (De Reu *et al.* 2011, 2438). As a result, the positive values of the TPI indicate that the cell is higher than its neighbourhood and inversely for negative values.

Landform reconstructions thus differ from slopes which are not computed through focal statistics.

A model of the landforms can be approximated through different reclassifications. Some studies employ a reclassification through manual thresholds, while others reclassify the TPI with standard deviation units (Weiss 2001; Nüsslein 2016b, 322). TPI is also the most used method of landform approximation, because of its very easy implementation in any GIS software. The choice of the radius size is very important and after several attempts with radii between 250m and 2500m, the radius of 1000m gave the best balance between precision and coherence. The resulting reclassified raster, usually with a set of 5 to 10 classes, can correspond – from the lowest value to the highest – to deeply incised valleys, valley bottoms, lower slopes, flat slopes, mid slopes, upper slopes, hilltops, canyons and other specific elements of the landscape, depending on the specific study area (fig.16).

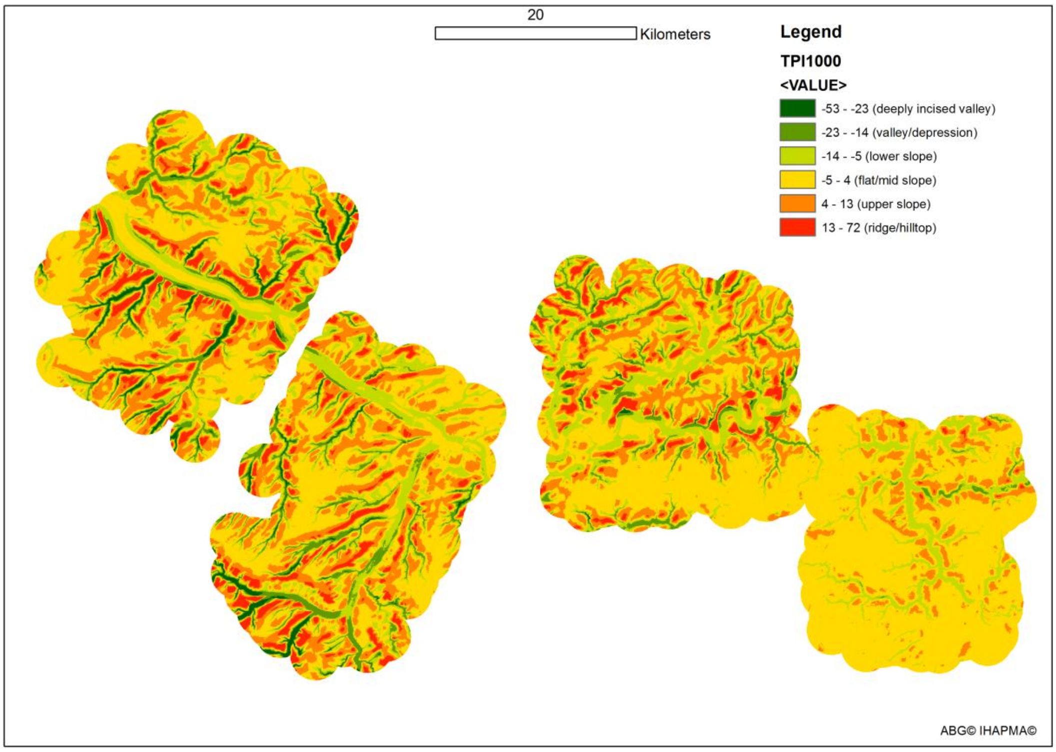


Figure 16: Topographic Position model, with a 1000m radius and six classes defined by standard deviation, IHAPMA© ABG©

The DEV method consists in the division of the TPI by the standard deviation of the DEM. The result cancels the local roughness around the cells and creates a more precise result. De Reu *et al.*'s studies show that TPI usually ends up with a very smooth landscape classification (2013, 45), where subtle changes in relatively flat areas are invisible, therefore creating disproportionately important areas for flat landforms. Conversely, DEV can create a much more detailed profile of even the flatter areas, generally producing analogous classes' areas after standard deviation reclassification, which correlate much better with other data, such as soil and geological maps or drainage classes (De Reu *et al.* 2013, 45). This higher representation and precision of classes is preferable in this modelling approach, as overrepresentation of any class would considerably diminish the predictive values. The comparison of the TPI and DEV models in this study area – computed with a radius of 1000m and reclassified by standard deviation – allows the same deductions as previously mentioned (fig. 17 and 18): in the fourth micro-region especially, the classes' precision is immensely improved through the use of DEV, as this region is characterised by very discrete elevation changes, virtually ignored by the TPI method. For this reason the DEV method only will be tested in a univariate model.

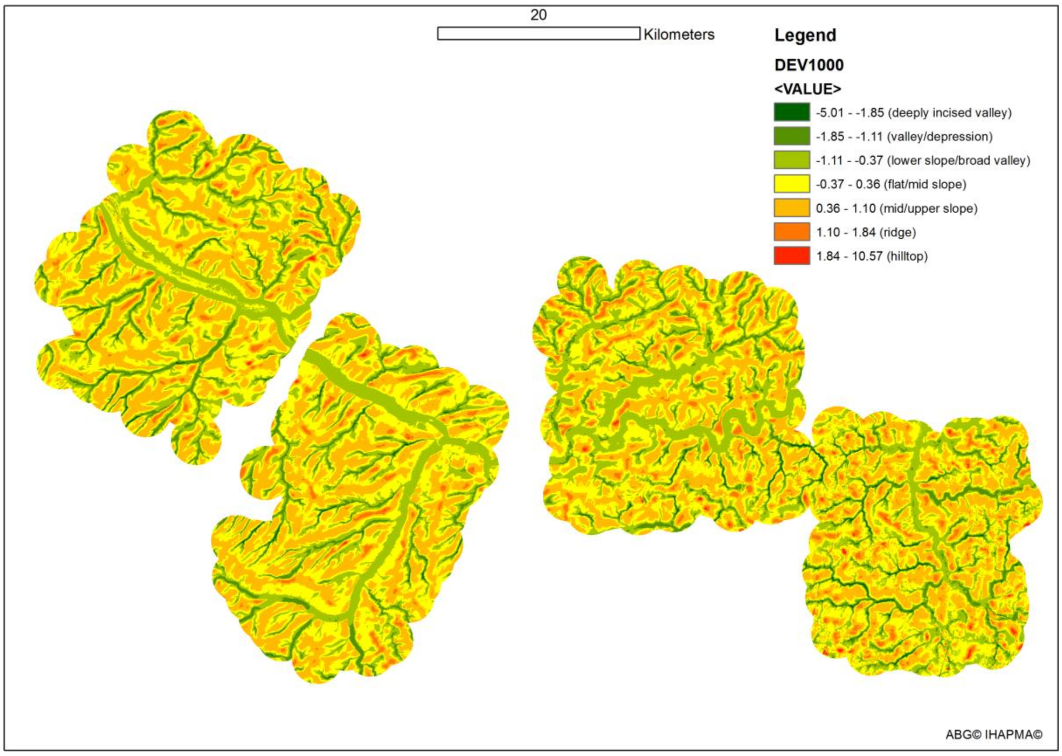


Figure 17: Deviation from mean elevation model, with a 1000m radius and seven classes defined by standard deviation, IHAPMA© ABG©

The DEV raster was further reclassified so as to represent the model of land suitability according to landforms (fig.18). Valleys and depressions were given a weight of 1, as they are the least suitable areas for settlement, where erosion is stronger in dry valleys and probably prone to flooding in active valleys, if not simply permanently under water. Lower and upper slopes, broad valleys and hilltops were given a weight of 2, as in this study area they do not actually constitute a real obstacle: this class contains gently sloping valleys where cattle could graze, or arable plateau summits. Finally, the flat-mid slopes and the plateaus' upper slopes and 'ridges' should be the most suitable areas for settlement, as they are generally well drained and arable: they therefore have a weight of 3.

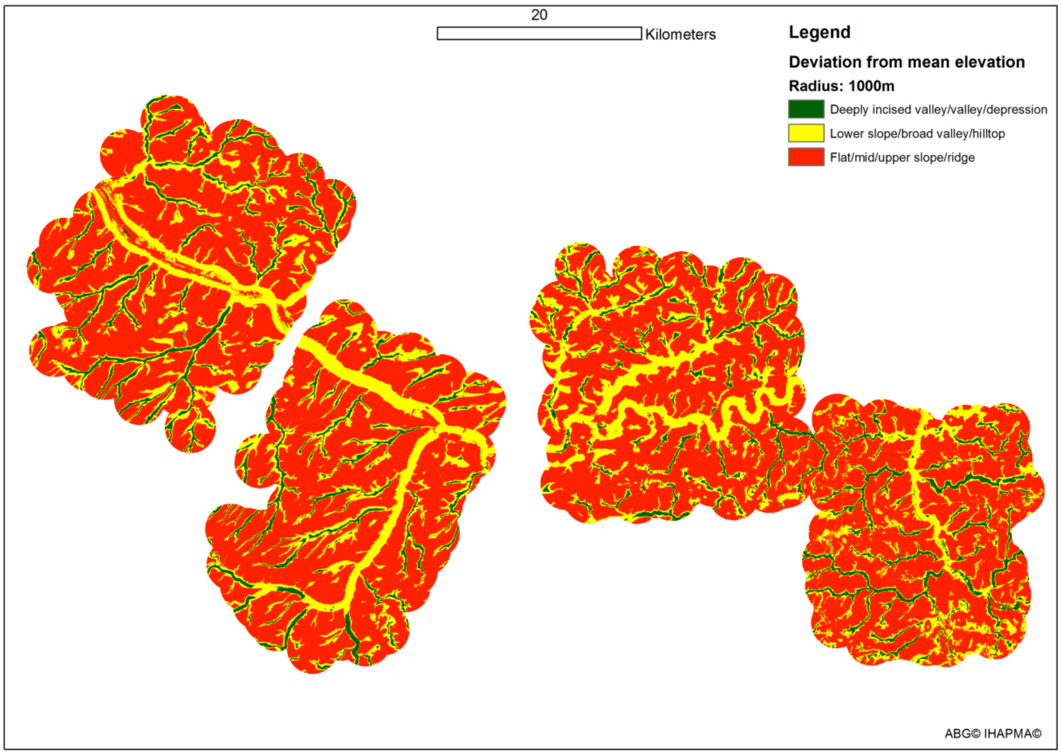


Figure 18: Univariate model of landform classification, based on the deviation from mean elevation in a radius of 1000m, IHAPMA© ABG©

The result of this reclassification (table 6), if it follows the precedent theory, is not entirely satisfying in that only the deepest valleys have a low weight value and the high weight area is very large.

Deviation from mean elevation class	**Weight**	**Observed sites**	**Expected sites**	P_o	**Area ratio** P_a	**Kvamme's gain** P_a/P_o	**Indicative value** P_o/P_a	**Relative gain** P_o-P_a	**Chi-square test**
Region 1									
Deep valley	1	6	14.97	0.03	0.07	-1.50	0.40	-0.04	0.00324
Hilltops/ lower slopes	2	34	45.96	0.16	0.22	-0.35	0.74	-0.06	
Flat/mid/ upper slopes	3	168	147.07	0.81	0.71	0.12	1.14	0.10	
Region 2									
Deep valley	1	10	10.91	0.06	0.06	-0.09	0.92	-0.01	0.78340
Hilltops/ lower slopes	2	42	38.34	0.25	0.22	0.09	1.10	0.02	
Flat/mid/ upper slopes	3	119	121.75	0.70	0.71	-0.02	0.98	-0.02	
Region 3									
Deep valley	1	5	13.02	0.02	0.06	-1.60	0.38	-0.04	0.00000
Hilltops/ lower slopes	2	26	52.76	0.13	0.26	-1.03	0.49	-0.13	
Flat/mid/ upper slopes	3	173	138.23	0.85	0.68	0.20	1.25	0.17	
Region 4									
Deep valley	1	14	20.70	0.06	0.09	-0.48	0.68	-0.03	0.00019
Hilltops/ lower slopes	2	28	50.34	0.12	0.21	-0.80	0.56	-0.09	
Flat/mid/ upper slopes	3	197	167.96	0.82	0.70	0.15	1.17	0.12	

Table 6: Results of the univariate model of the landform classification based on deviation from mean elevation, ABG© IHAPMA©

First of all, the results show a good Chi-square test in all regions but the second, which shows a near homogeneity between surface ratio and site ratio. The other regions behave with an important heterogeneity in regard to the expected values. The discrepancies lie between the low and medium predictive classes, which are very close in predictive value, or even inversed in the fourth region. In any case, that the very vast high prediction class indeed has slightly positive predictive value should allow further refining of the results through combination with other factors. The main issue to be solved should be to restrict the size of the high prediction area and the next variable should help in that direction.

3.2.5 Distance to water

As none of the studied micro-regions is in the vicinity of the North Sea, only active rivers will be taken into consideration in this variable. Using absolute distance/euclidian distance as a set of buffer areas around the rivers would be too coarse an approximation of the actual distance of water – especially considering the relatively low resolution of the vectorised rivers – which is why a path distance was rather created for this purpose. This method is generally employed in the modelling of least cost paths, i.e. for creating paths from one point source to one or several point destinations (Herzog 2010; Güimil-Fariña *et al.* 2015; Lock *et al.* 2010; Verhagen and Jeneson 2012; Verhagen *et al.* 2014). Though, it is also possible to model the relative distance from segments to any point of the environment. This method of analysis is related to what Lock and Harris called a 'sphere of influence', employed as a proxy for accessibility (Harris *et al.* 2006, 52–53).

This procedure consists in computing for every cell the cumulative minimal cost from the closest segment section. This cost is dependent on slope and elevation and the Tobler cost function was implemented for this model (Tobler 1993): it calculates the cost of each slope value so as to correlate with the actual effort needed for a hiker to travel such area. Thus, the minimal cost is reached for a slope of -2.85°, while negative and positive extremes have an exponentially higher cost. The unit of the Tobler cost function is the hour per meter, which means that the cumulative cost's unit is measured in hours, although this is an indicative value rather than a verifiable fact.

The aim of this model is double: on the one hand attractive areas will be built relatively close to rivers, and on the other hand a repulsive distance class will be created very close to the rivers (less than a minute of cost), indicating areas which could temporarily or permanently be under water. This class is an approximation of the past landscape and only proper geomorphological studies might shed light on the actual behaviour of each river. The cells with a cumulative cost above one hour from a river are also included in this low probability area. The two other classes are less prone to debate, as they depend on general and stable trends in the topography. The high probability class is therefore the second closest to the river, up to half an hour of cost and the medium probability class is above half an hour of cost but under an hour (fig.19).

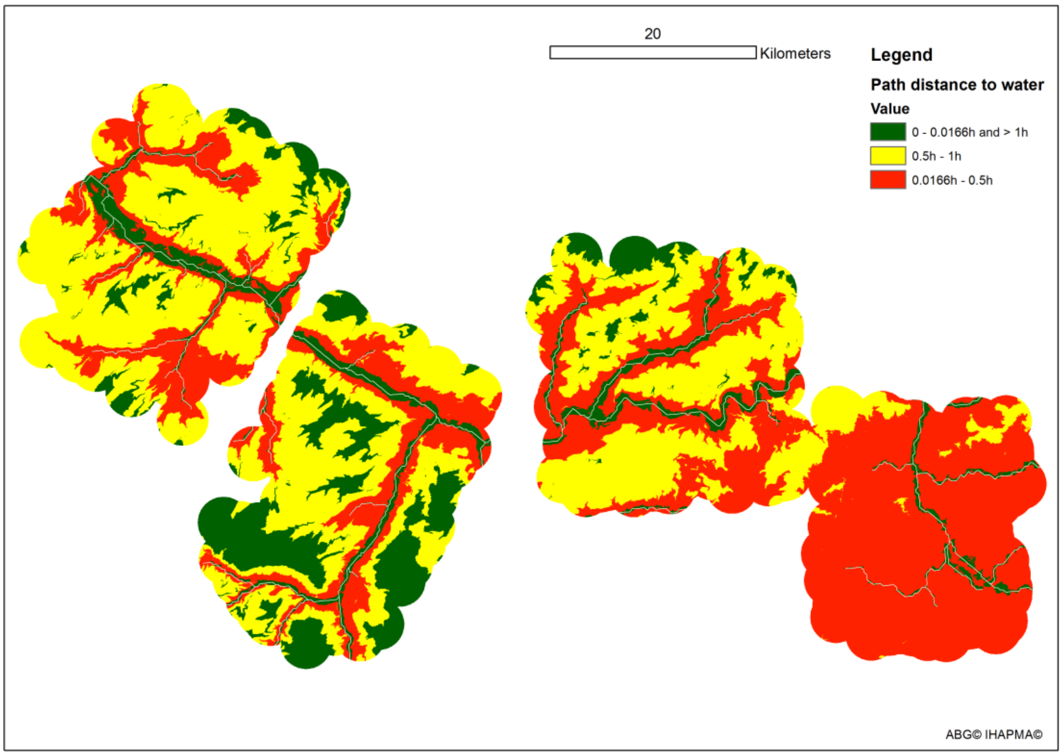


Figure 19: Univariate model of path distance (Tobler's cost function) to river courses, IHAPMA© ABG©

The results of the model (table 7) indicate that the precedent theory on site location was wrong. Indeed, if apart from the second region all regions display a very good heterogeneity between actual and expected sites, only the fourth region seems to have the highest predictive value in an intermediate location between 0.01666 hours and half an hour cost away from rivers. Even the results of this fourth region should be put in doubt, as the low and medium probability classes only make up for 6% of the region's area. The medium range of probability – concerning areas between half an hour and an hour cost away from rivers – seems to be the preferred site location in the three non-random models. Are sites preferably located as far as possible from river courses? Or does it simply mean that these areas constitute the fertile plateaus? The areas further than one hour to rivers do not display such settlement density, which means that this intermediary position in the path distance to water is indeed the preferred site location.

The homogeneity of the second region is due to an increased proportion of sites in the low probability area. Firstly, this can be explained by the large proportion of the region it occupies through areas a distance more than one hour to rivers, which are in fact well settled. Secondly, the local rivers of the Selle and the Evoissons display slightly more sites in their immediate vicinity. The site of Frémontiers indeed showcases built embankments and other storage facilities on the Evoissons, which indicates that for some specific areas, site location is preferable near well-contained rivers.

Distance class (hours)	**Weight**	**Observed sites**	**Expected sites**	P_o	**Area ratio** P_a	**Kvamme's gain** P_a/P_o	**Indicative value** P_o/P_a	**Relative gain** P_o-P_a	**Chi-square test**
Region 1									
0 - 0.0166 / >1	1	11	27.97	0.05	0.13	-1.54	0.39	-0.08	0.00154
0.5 - 1	2	145	126.76	0.70	0.61	0.13	1.14	0.09	
0.0166 - 0.5	3	52	53.27	0.25	0.26	-0.02	0.98	-0.01	
Region 2									
0 - 0.0166 / >1	1	51	57.58	0.30	0.34	-0.13	0.89	-0.04	0.34634
0.5 - 1	2	84	74.65	0.49	0.44	0.11	1.13	0.05	
0.0166 - 0.5	3	36	38.77	0.21	0.23	-0.08	0.93	-0.02	
Region 3									
0 - 0.0166 / >1	1	9	27.01	0.04	0.13	-2.00	0.33	-0.09	0.00006
0.5 - 1	2	128	101.89	0.63	0.50	0.20	1.26	0.13	
0.0166 - 0.5	3	67	75.10	0.33	0.37	-0.12	0.89	-0.04	
Region 4									
0 - 0.0166 / >1	1	3	10.25	0.01	0.04	-2.42	0.29	-0.03	0.00470
0.5 - 1	2	11	20.32	0.05	0.09	-0.85	0.54	-0.04	
0.0166 - 0.5	3	225	208.43	0.94	0.87	0.07	1.08	0.07	

Table 7: Results of the univariate model of path distance (Tobler's cost function) to river courses, ABG© IHAPMA©

In order to be combined with other variables in the next multivariate weighted model, the distance to water classification will therefore be changed, acknowledging this new understanding of site location preferences. Thus, the low probability area will only contain sites which are closer than a minute-cost (0.0166 hour) to the rivers, while the medium probability area will contain sites which are between a minute and half an hour from rivers, as well as sites beyond one hour. Finally, the high probability class will encompass a slightly expanded area between half an hour and an hour cost from rivers. Here are the results of this new model (fig.20 and table 8):

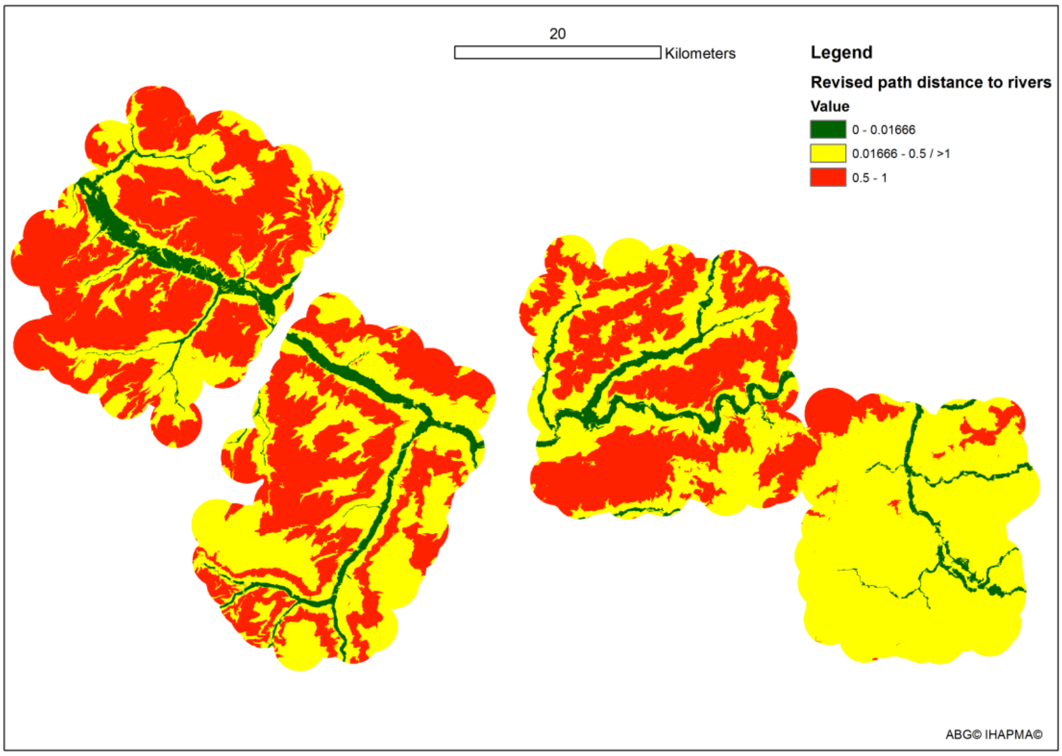


Figure 20: Revised univariate model of path distance to rivers, IHAPMA© ABG©

The performance of the model is slightly improved in terms of Chi-square and predictive values: regions 1 and 3 now display values which correlate with the precedent theory. Although the randomness of region 2 could not be changed, it is now even more apparent that the immediate vicinity of rivers has a different role in this area. The values in region 4 were simply inversed, but the relative smoothness of the landscape in this area creates unrealistic surfaces for the distance path: the medium area is supposed to represent up to half an hour distance from rivers, but it is clear that the modelled surfaces go well beyond this limit.

Distance class (hours)	Weight	Observed sites	Expected sites	P_o	Area ratio P_a	Kvamme's gain P_a/P_o	Indicative value P_o/P_a	Relative gain P_o-P_a	Chi-square test
Region 1									
0 - 0.0166	1	7	13.51	0.03	0.06	-0.93	0.52	-0.03	0.02029
0.0166 - 0.5 / >1	2	56	67.73	0.27	0.33	-0.21	0.83	-0.06	
0.5 - 1	3	145	126.76	0.70	0.61	0.13	1.14	0.09	
Region 2									
0 - 0.0166	1	11	10.22	0.06	0.06	0.07	1.08	0.00	0.29776
0.0166 - 0.5 / >1	2	76	86.13	0.44	0.50	-0.13	0.88	-0.06	
0.5 - 1	3	84	74.65	0.49	0.44	0.11	1.13	0.05	
Region 3									
0 - 0.0166	1	3	14.05	0.01	0.07	-3.68	0.21	-0.05	0.00013
0.0166 - 0.5 / >1	2	73	88.06	0.36	0.43	-0.21	0.83	-0.07	
0.5 - 1	3	128	101.89	0.63	0.50	0.20	1.26	0.13	
Region 4									
0 - 0.0166	1	3	10.25	0.01	0.04	-2.42	0.29	-0.03	0.00470
0.0166 - 0.5 / >1	2	225	208.43	0.94	0.87	0.07	1.08	0.07	
0.5 - 1	3	11	20.32	0.05	0.09	-0.85	0.54	-0.04	

Table 8: Results of the revised univariate model of path distance to rivers, ABG© IHAPMA©

To conclude with the analysis of this variable, very sharp differences can be seen in the behaviour of this model in the different micro-regions. If it seems certain that this variable is correlated with site location, the part it plays might not be very important, or has not been approached in the proper way. By combining this factor with others of preceding relevance, a more accurate and better-performing model of the physical landscape's influence on site location can be built.

3.2.6 Weighted multivariate model of the physical landscape

Now that each available variable concerning the physical landscape has been analysed individually, a weighted multivariate model will be constructed in order to verify this theory on site location. It has been noted that aspect and solar radiation could not be taken into account at all because of their high homogeneity between expectations and site location; they shall therefore not be combined in the next model. Strong doubts towards the path distance to rivers have also been emitted, so its role will be weighted down. The best individual predictive values were reached by the DEV classification of landforms, while slopes are not far behind. Therefore this formal theory is proposed in the form of an equation:

> (DEV X 1.5 + Slopes X 1 + Distance to water X 0.5) / 3 = Positive site location correlation in attractive classes

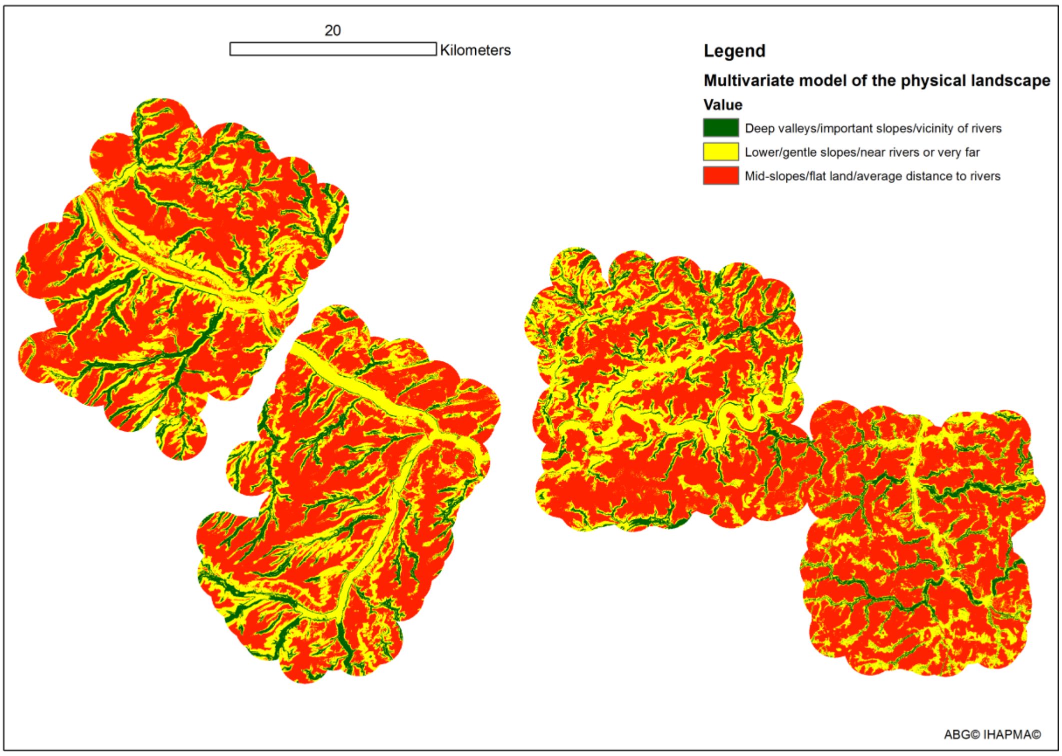


Figure 21: Multivariate model of the physical landscape (DEV, slopes and distance to water), IHAPMA© ABG©

The resulting raster (fig. 21) and its results (table 9) show the results of the reclassification of categorical weights according to equal interval from 1 to 3, therefore from 1 to 1.66, 1.66 to 2.33 and 2.33 to 3. The micro-regions as well as the combination of all regions are presented so that local and general trends might be compared.

Physical landscape	Predictive class	Observed sites	Expected sites	P_o	Area ratio P_a	Kvamme's gain P_a/P_o	Indicative value P_o/P_a	Relative gain P_o-P_a	Chi-square test
All regions									
	1	50	86.09	0.06	0.10	-0.72	0.58	-0.04	0.00000
	2	172	264.23	0.21	0.32	-0.54	0.65	-0.11	
	3	600	471.69	0.73	0.57	0.21	1.27	0.16	
Region 1									
	1	14	24.39	0.07	0.12	-0.74	0.57	-0.05	0.00001
	2	44	67.94	0.21	0.33	-0.54	0.65	-0.12	
	3	150	115.66	0.72	0.56	0.23	1.30	0.17	
Region 2									
	1	16	19.84	0.09	0.12	-0.24	0.81	-0.02	0.02553
	2	44	57.76	0.26	0.34	-0.31	0.76	-0.08	
	3	111	93.40	0.65	0.55	0.16	1.19	0.10	
Region 3									
	1	9	19.54	0.04	0.10	-1.17	0.46	-0.05	0.00000
	2	38	73.47	0.19	0.36	-0.93	0.52	-0.17	
	3	157	110.99	0.77	0.54	0.29	1.41	0.23	
Region 4									
	1	15	20.24	0.06	0.08	-0.35	0.74	-0.02	0.00002
	2	31	58.89	0.13	0.25	-0.90	0.53	-0.12	
	3	193	159.87	0.81	0.67	0.17	1.21	0.14	

Table 9: Results of the multivariate model of the physical landscape (DEV, slope and distance to rivers), ABG© IHAPMA©

As can be seen, the general results are satisfying, in that all show a very good heterogeneity between expected and actual site proportions and the most predictive area is always the one that was expected to be. Taking all regions into consideration offers a relatively good efficiency, with low, medium and high probability classes being discriminated by well-distributed predictive values. Nonetheless, some criticism should be given towards this model. The best predictive values are reached in the third region, but this gain is not very important (not more than 41% or a relative gain of 0.21). In all regions the medium and low predictive classes are very close to each other in terms of site frequency, sometimes still reversed and the low predictive class is generally too predictive, its relative value always being above the medium class' one. The general predictive accuracy of the model is therefore average.

Despite these remarks, this model still provides precision over accuracy. Even with the combination of three very different rasters, the topographical landforms and the highest slopes are still visible and play an important part in the model's behaviour. Many predictive models employ focal statistics during the construction and the test of the model, but when applying it to this case study, even with a relatively small radius (250m), most details were erased from the map. At the same time, the medium predictive class is inflated, taking some predictive value from the former high predictive class and leaving an even smaller low predictive class behind. Therefore, no attempt at smoothing down the result through focal statistics will be carried out in the following chapters.

Now the model's behaviour in different sub-periods will be evaluated. This overview will be restricted to two sub-periods: the first half of the second century – which is characterised by the highest settlement intensity in the Somme – and the first half of the fourth century, a period of heavy decline, massive abandonment, but with hopefully enough sites to still represent some settlement patterns, which is no more the case in the second half of the fourth century and even more so in the fifth century. A table for each sub-period is given below, followed by a brief overview of the changes which can be highlighted from the results. The tables include for each predictive index its deviation from the former model.

In the first half of the second century (table 10), only the first and third regions display a sufficiently low Chi-square test result; all other regions are therefore affected by a certain random character. The low predictive class is actually more predictive than the medium one in all regions but the first, which indicates a low correlation between site location during this period and the supposedly limiting factors of the physical environment, although there are more sites in the high predictive class. Deviation from the global model is therefore noticeable and exacerbates the relative homogeneity of site location during this period.

First half of the 2nd century and the physical landscape	Predictive class	Observed sites	Expected	P_o	Area ratio P_a	Kvamme's gain	Indicative value	Relative gain	Chi-square test
Region 1									
	1	4	9.03	0.05	0.12	-1.26	0.44	-0.07	0.04330
	2	20	25.15	0.26	0.33	-0.26	0.80	-0.07	
	3	53	42.82	0.69	0.56	0.19	1.24	0.13	
Region 2									
	1	8	9.16	0.10	0.12	-0.15	0.87	-0.01	0.19689
	2	20	26.68	0.25	0.34	-0.33	0.75	-0.08	
	3	51	43.15	0.65	0.55	0.15	1.18	0.10	
Region 3									
	1	4	7.28	0.05	0.10	-0.82	0.55	-0.04	0.00010
	2	12	27.37	0.16	0.36	-1.28	0.44	-0.20	
	3	60	41.35	0.79	0.54	0.31	1.45	0.25	
Region 4									
	1	5	5.42	0.08	0.08	-0.08	0.92	-0.01	0.12584
	2	9	15.77	0.14	0.25	-0.75	0.57	-0.11	
	3	50	42.81	0.78	0.67	0.14	1.17	0.11	

Table 10: Results of the multivariate model of the physical landscape (DEV, slope and distance to rivers) during the first half of the second century, ABG© IHAPMA©

For the first half of the fourth century (table 11), the results are quite contradictory. Indeed, the low predictive class can be rather predictive according to the indicative values in all regions apart from the third, and all indices drop in the medium class. It could be interpreted that there are differential survival rates of sites located in the valleys and the plains, whereas average lands were massively abandoned, but we must be doubtful of any result based on a sample of so few sites.

First half of the 4th century and the physical landscape	Predictive class	Observed sites	Expected sites	P_o	Area ratio P_a	Kvamme's gain	Indicative value	Relative gain	Chi-square test
Region 1									
	1	2	2.46	0.10	0.12	-0.23 (+0.51)	0.81 (+0.24)	-0.02 (+0.03)	0.32881
	2	4	6.86	0.19	0.33	-0.71 (-0.17)	0.58 (-0.07)	-0.14 (-0.02)	
	3	15	11.68	0.71	0.56	0.22 (-0.01)	1.28 (-0.02)	0.16 (-0.01)	
Region 2									
	1	3	3.83	0.09	0.12	-0.28 (-0.04)	0.78 (-0.03)	-0.03 (-0.01)	0.04355
	2	5	11.15	0.15	0.34	-1.23 (-0.92)	0.45 (-0.31)	-0.19 (-0.11)	
	3	25	18.03	0.76	0.55	0.28 (+0.12)	1.39 (+0.20)	0.21 (+0.11)	
Region 3									
	1	0	3.26	0.00	0.10	0.00 (+1.17)	0.00 (-0.46)	-0.10 (-0.05)	0.00117
	2	5	12.24	0.15	0.36	-1.45 (-0.52)	0.41 (-0.11)	-0.21 (-0.04)	
	3	29	18.50	0.85	0.54	0.36 (+0.07)	1.57 (+0.16)	0.31 (+0.08)	
Region 4									
	1	2	2.29	0.07	0.08	-0.14 (+0.21)	0.87 (+0.13)	-0.01 (+0.01)	0.23451
	2	3	6.65	0.11	0.25	-1.22 (-0.32)	0.45 (-0.08)	-0.14 (-0.02)	
	3	22	18.06	0.81	0.67	0.18 (+0.01)	1.22 (+0.01)	0.15 (+0.01)	

Table 11: Results of the multivariate model of the physical landscape (DEV, slope and distance to rivers) during the first half of the fourth century, ABG© IHAPMA©

3.2.7 General conclusions on the influence of the physical landscape

It can be concluded that if the starting hypothesis is not disproved by the model's results, one must remain vigilant not to expect too much deterministic limitation from the landscape, as all of its features were settled, if in different frequencies. Very few places were truly limiting for Roman rural settlement location – especially in the fourth region – and this model's accuracy is reduced by the very large proportion of very attractive land. All the different types of landscape were settled at one point or another during the Roman period, be it in the valleys, the plateaus or the plains and if the preference clearly goes towards flat and mid-slopes with a good access to water and few slopes, nothing clearly prevented settlement in the other areas.

Trying to ‘fix’ the variables so that they correlate closer and closer to the site location, even if it creates a compelling predictive model, is not the present aim: it may simply be that some categories of the landscape are not influencing site location or only in some places. Solely through geoarchaeological reconstructions of soil distributions and hydrological basins could a better starting point be modelled. Nevertheless, the current model of the physical landscape will provide a precise background upon which the social factors can be modelled in the following chapter.

3.3 Model of the socio-economic variables

In this chapter, the following question will be inquired: Is the physical environment less influential than socio-economic parameters in site location and which of the latter play the most important roles? One hypothesis is that rural site location may indeed depend mainly on the accessibility of several socio-economic focal points in the neighbouring rural settlements of the human environment. In theory, proximity to Roman roads, main cities and secondary foci should produce highly predictive classes.

Specific variables will be created and evaluated towards a reconstruction of a socio-economic environment which could have influenced the location of rural settlements. For this purpose, bivariate models will be built, using one socio-economic variable at a time and the final model of physical landscape, before its final reclassification (with a wide range of values between 1 and 3). Thereafter, the predictive values will also include the deviation in terms of positive or negative gains to the model of the physical landscape. That way, each variable's influence will appear clearly in maps and predictive values. In the end, analogous and pertinent variables will be combined in a weighted multivariate model, which in turn will be combined with the final model of the physical landscape. Because of computational power limits and larger extents, the new variables can only have a cell resolution of 75m.

The first variables pertain to path distances from specific sites or Roman roads and are therefore additive and compatible. The other variables which will be created relate to openness and visibility – which are DEM-derived – and land-use heritage, whose methodology should be set apart from the other variables.

3.3.1 Euclidian distance to roads

This first variable is very straightforward: a raster of Euclidian distance is created with the reconstructed Roman network of main roads, which are valuable vectors of interconnection and economic development. This raster is then reclassified in categories of 1000m distance, 2000m and above 2000m and combined with the physical landscape. The closer to a road, the higher the probability of site location. As the road network is stable throughout the whole period studied and often well beyond, the evaluation of the model is applied without any concern towards the chronology.

The results of the model display massive visual and statistical changes (fig. 22 and table 12). Indeed, the high predictive class is only located less than 2000m from a Roman road, which reduces considerably its area and increases the surface of medium and low predictive classes. The roads pass indiscriminately straight through diverse landscape classes, increasing site location probability in areas where the physical environment is less attractive than elsewhere. By comparing each predictive classes, they now are much more balanced in terms of area, apart from the first region where only a small section of the *via* towards Boulogne-sur-Mer is currently known. Because of this lack of representation of the road network in this region, its behaviour is contrary to the theory. In the other regions, in terms of predictive gains, this model showcases a large improvement upon the physical landscape alone: the medium prediction area is now much more predictive, neighbouring the indicative value of 1, while the low prediction area has lowered predictive values. The region which has the best performance regarding the theory is the third, where predictive values are very well segregated, with the maximum gain in the high predictive class until now. The much increased medium predictive values in the fourth region, compared to its high predictive class, are not refuting the theory either, as the medium prediction class for the distance to roads still represents areas very close to roads. It can be confidently concluded that – notwithstanding the first region – the Euclidian distance to roads improves the model considerably, globally agreeing with the precedent theory.

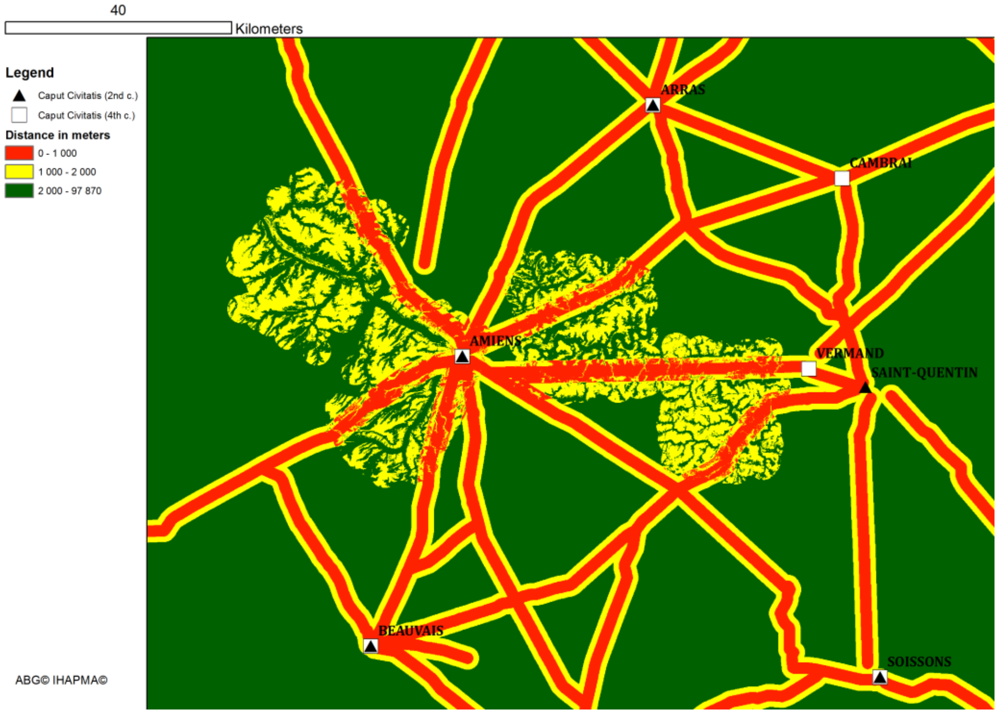


Figure 22: Bivariate model of the Euclidian distance to roads and the physical landscape, IHAPMA© ABG©

Euclidian distance to roads	Predictive class	Observed sites	Expected	P_o	Area ratio P_a	Kvamme's gain	Indicative value	Relative gain	Chi-square test
Region 1									
	2000-97870 m	54	84.96	0.26	0.41	-0.57 (+0.17)	0.64 (+0.07)	-0.15 (-0.10)	0.00001
	1000-2000m	143	109.22	0.69	0.53	0.24 (+0.78)	1.31 (+0.66)	0.16 (+0.28)	
	0-1000 m	11	13.82	0.05	0.07	-0.26 (-0.49)	0.80 (-0.50)	-0.01 (-0.18)	
Region 2									
	2000-97870 m	40	49.54	0.23	0.29	-0.24 (0)	0.81 (0)	-0.06 (-0.04)	0.13489
	1000-2000m	66	67.25	0.39	0.39	-0.02 (+0.29)	0.98 (+0.22)	-0.01 (+0.07)	
	0-1000 m	65	54.21	0.38	0.32	0.17 (+0.01)	1.20 (+0.01)	0.06 (-0.04)	
Region 3									
	2000-97870 m	27	66.59	0.13	0.33	-1.47 (-0.30)	0.41 (-0.05)	-0.19 (-0.14)	0.00000
	1000-2000m	93	86.44	0.46	0.42	0.07 (+1.00)	1.08 (+0.56)	0.03 (+0.20)	
	0-1000 m	84	50.97	0.41	0.25	0.39 (+0.10)	1.65 (+0.24)	0.16 (-0.07)	
Region 4									
	2000-97870 m	24	54.61	0.10	0.23	-1.28 (-0.93)	0.44 (-0.30)	-0.13 (-0.11)	0.00001
	1000-2000m	155	130.91	0.65	0.55	0.16 (+1.06)	1.18 (+0.65)	0.10 (+0.22)	
	0-1000 m	60	53.48	0.25	0.22	0.11 (-0.06)	1.12 (-0.09)	0.03 (-0.11)	

Table 12: Results of the bivariate model of the Euclidian distance to Roman roads and the physical landscape, with value deviations from the model of physical landscape alone, ABG© IHAPMA©

3.3.2 Path distance from *capita civitatum*

The easiest and some of the most influential variables input of a predictive model are the archaeological remains. The Roman Empire as a whole was a very interconnected place, economically, culturally and socially. The *capita civitatum* in particular played a major role in every region in the Empire: they governed and centralised the *civitas* and rural settlements, and their owners and secondary agglomerations heavily depended on this focal nexus.

Two different path distance rasters will be created, so as to evaluate the behaviour of the model from a global perspective and in two different periods: the first half of the second century and the first half of the fourth century. The analysis of these two periods will allow us to look at the changes in the influence of the *capita* during the optimum period of the land-use by rural settlements. Concerning the Late Roman period and its much lesser demographic density, the model will also help understand if site location is influenced by this factor, and furthermore if the displacement of the *caput civitatis viromanduorum* from Saint-Quentin to Vermand and of the *caput civitatis nerviorum* from Bavay to Cambrai, both now further west and therefore very close to the fourth micro-region, had a negative impact or not.

Therefore, two of the models will require specific site assemblages which have a chronological classification. This represents a huge loss of information: out of the 822 rural settlements in the micro-regions, 296 are proved to be occupied during the first half of the second century and only 115 in the fourth century, among which only 18 did not exist already in the first half of the second century. For the global model – which employs dated and undated sites alike – the model evaluated is the second century one, as it represents the vast majority of the Roman period's administrative setting, as well as an immense proportion of all sites, for only a handful were created during the Late Roman period.

Using Euclidian distance, as has been explained earlier, would create an unrealistic model which would contradict heavily with the model of the physical landscape with which it is supposed to be later combined. Using the same approach as earlier for the distance to water, cost surfaces were modelled with Tobler's hiking algorithm and then reclassified in categories of maximum 1.75 hours cost, maximum 2.5 hours cost and above 2.5 hours cost, with the closer category having the highest predictive expectation. These values were manually defined so that each micro-region contained at least a small proportion of all three categories.

Because of the heavy computational power needed for this process and also because its geographical extent encompasses a much larger area, the DEM from which the slope raster and the path distance are computed is the 75m-resolution IGN RGE ALTI©. For a clearer understanding of the model, the maps given below (fig.23 and 24) represent a larger extent than the one used for the statistical evaluation which is based on the micro-regions' mask. The same geographical problems encountered earlier with path distance to rivers are found again in the model: hour-costs are disproportionately low compared to actual distance when the path distance meets the very gentle slopes of major valleys. In this case, this distortion works in favour of the theory in the sense that it emulates accessibility from fluvial transport in the Somme and its major tributaries.

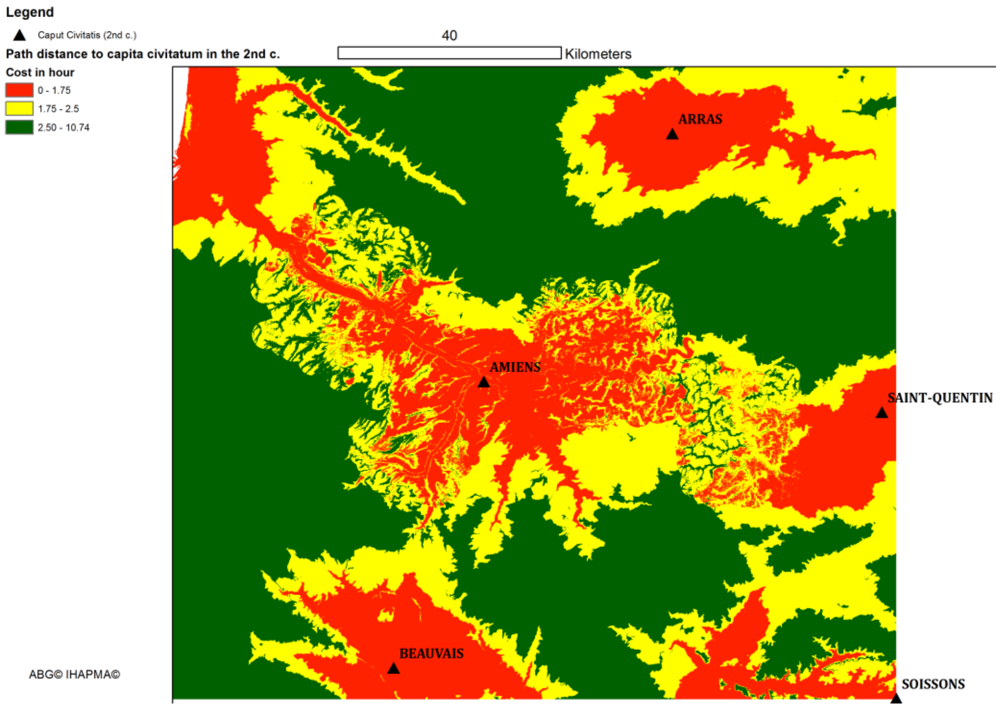


Figure 23: Bivariate model of the path distance to *capita civitatum* and the physical landscape in the first half of the second century, IHAPMA© ABG©

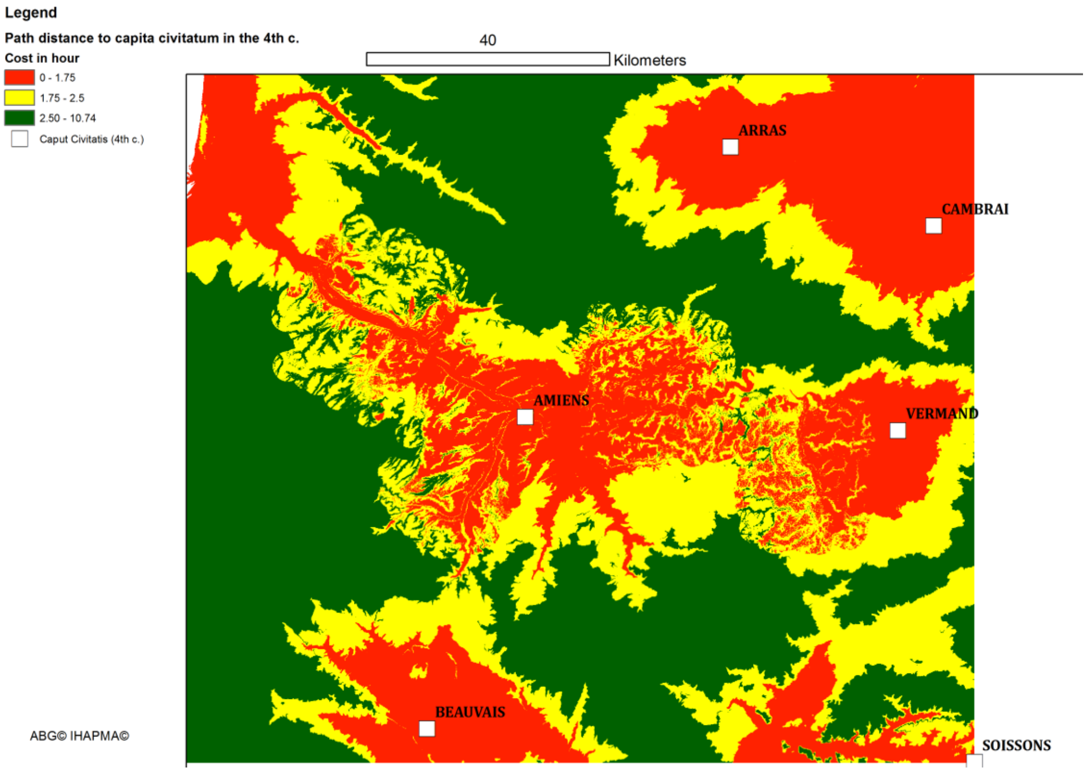


Figure 24: Bivariate model of the path distance to capita civitatum and the physical landscape during the first half of the fourth century, IHAPMA© ABG©

The results of the global model are rather in accordance with the theory (table 13): all Chi-square results are very good and each class is well discriminated, apart from the fourth region where the high prediction class is not predictive enough, which may be explained by the massive increase in predictive values for the medium class. Because of the large distance values put in the reclassification of the path distance, much fewer areas are in the low prediction class, the latter nevertheless displaying a very low predictive value in all classes. There is therefore a strong correlation between inaccessibility from cities and low predictive value, which might imply an influence on site location. Though, areas which are very accessible from cities do not display much higher predictive values, which imply that site location is relatively homogeneous below a certain accessibility value under which it would be influenced by other factors.

Path distance to capita civitatum (All sites)	Predictive class	Observed sites	Expected	P_o	Area ratio P_a	Kvamme's gain	Indicative value	Relative gain	Chi-square test
Region 1									
	2.25-10 h	21	45.80	0.10	0.22	-1.18 (-0.44)	0.46 (-0.11)	-0.12 (-0.07)	0.00012
	1.75-2.25 h	120	108.87	0.58	0.52	0.09 (+0.63)	1.10 (+0.45)	0.05 (+0.17)	
	0-1.25 h	67	53.32	0.32	0.26	0.20 (-0.03)	1.26 (-0.04)	0.07 (-0.10)	
Region 2									
	2.25-10 h	3	12.07	0.02	0.07	-3.02 (-2.78)	0.25 (-0.56)	-0.05 (-0.03)	0.00396
	1.75-2.25 h	45	53.46	0.26	0.31	-0.19 (+0.12)	0.84 (+0.08)	-0.05 (+0.03)	
	0-1.25 h	123	105.47	0.72	0.62	0.14 (-0.02)	1.17 (-0.02)	0.10 (0.00)	
Region 3									
	2.25-10 h	4	14.31	0.02	0.07	-2.58 (-1.41)	0.28 (-0.18)	-0.05 (0.00)	0.00002
	1.75-2.25 h	48	68.63	0.24	0.34	-0.43 (+0.50)	0.70 (+0.18)	-0.10 (+0.07)	
	0-1.25 h	152	121.06	0.75	0.59	0.20 (-0.09)	1.26 (-0.15)	0.15 (-0.08)	
Region 4									
	2.25-10 h	15	27.06	0.06	0.11	-0.80 (-0.45)	0.55 (-0.19)	-0.05 (-0.03)	0.03679
	1.75-2.25 h	158	144.76	0.66	0.61	0.08 (+0.98)	1.09 (+0.56)	0.06 (+0.18)	
	0-1.25 h	66	67.18	0.28	0.28	-0.02 (-0.19)	0.98 (-0.23)	0.00 (-0.14)	

Table 13: Results of the bivariate model of the path distance to *capita civitatum* and the physical landscape – regardless of chronology and on the model of the Early Empire's urban setting – with value deviations from the model of physical landscape alone, ABG© IHAPMA©

The evaluation of the model in the first half of the second century concurs with the preceding analysis (table 14), with the exact same trends visible in every class and region, maybe even more strongly discriminated. The Chi-square test shows a high randomness though, mainly due to the much reduced number of sites. The same can be said of the model for the first half of the fourth century (fig. 24 and table 15), with even less accuracy or even no site at all, but surprisingly the same statistical trends again. The move of *caput* from Saint-Quentin to Vermand does not seem to have much influence on site location in the fourth region. On the subject of distance to *capita civitatum*, it can therefore be concluded that all periods show similar trends in site location, with inaccessibility being more influential than a very high accessibility. One must be careful not to overanalyse the results for specific periods, as their randomness index is very high, which is due to the very small number of well-dated sites and therefore too small number of site location in all classes. It is expected that the influence of secondary agglomerations should actually be larger and the results more balanced.

Path distance to *capita civitatum* (2nd century)	**Predictive class**	**Observed sites**	**Expected**	**P_o**	**Area ratio P_a**	**Kvamme's gain**	**Indicative value**	**Relative gain**	**Chi-square test**
Region 1									
	2.25-10 h	3	15.64	0.04	0.22	-4.21	0.19	-0.18	0.00003
	1.75-2.25 h	36	37.16	0.51	0.52	-0.03	0.97	-0.02	
	0-1.25 h	32	18.20	0.45	0.26	0.43	1.76	0.19	
Region 2									
	2.25-10 h	1	4.52	0.02	0.07	-3.52	0.22	-0.05	0.11832
	1.75-2.25 h	17	20.01	0.27	0.31	-0.18	0.85	-0.05	
	0-1.25 h	46	39.48	0.72	0.62	0.14	1.17	0.10	
Region 3									
	2.25-10 h	1	4.56	0.02	0.07	-3.56	0.22	-0.05	0.12080
	1.75-2.25 h	19	21.87	0.29	0.34	-0.15	0.87	-0.04	
	0-1.25 h	45	38.57	0.69	0.59	0.14	1.17	0.10	
Region 4									
	2.25-10 h	4	5.43	0.08	0.11	-0.36	0.74	-0.03	0.65758
	1.75-2.25 h	32	29.07	0.67	0.61	0.09	1.10	0.06	
	0-1.25 h	12	13.49	0.25	0.28	-0.12	0.89	-0.03	

Table 14: Results of the bivariate model of the path distance to capita civitatum and the physical landscape during the first half of the second century, ABG© IHAPMA©

Path distance to *capita civitatum* (4th c.)	**Predictive class**	**Observed sites**	**Expected**	**P_o**	**Area ratio P_a**	**Kvamme's gain**	**Indicative value**	**Relative gain**	**Chi-square test**
Region 1									
	2.25-10 h	0	4.62	0.00	0.22	0.00	0.00	-0.22	0.02806
	1.75-2.25 h	12	10.99	0.57	0.52	0.08	1.09	0.05	
	0-1.25 h	9	5.38	0.43	0.26	0.40	1.67	0.17	
Region 2									
	2.25-10 h	0	2.33	0.00	0.07	0.00	0.00	-0.07	0.08371
	1.75-2.25 h	7	10.32	0.21	0.31	-0.47	0.68	-0.10	
	0-1.25 h	26	20.35	0.79	0.62	0.22	1.28	0.17	
Region 3									
	2.25-10 h	0	2.39	0.00	0.07	0.00	0.00	-0.07	0.00720
	1.75-2.25 h	5	11.44	0.15	0.34	-1.46	0.41	-0.21	
	0-1.25 h	29	20.18	0.85	0.59	0.25	1.34	0.21	
Region 4									
	2.25-10 h	0	1.06	0.00	0.04	0.00	0.00	-0.04	0.15232
	1.75-2.25 h	15	10.49	0.56	0.39	0.27	1.37	0.15	
	0-1.25 h	12	15.45	0.44	0.57	-0.34	0.75	-0.15	

Table 15: Results of the bivariate model of the path distance to *capita civitatum* and the physical landscape in the first half of the fourth century, ABG© IHAPMA©

3.3.3 Path distance from local foci

The secondary agglomerations/villages/*vici* play the second role in the economic network of the *civitas*, notwithstanding forts and military settlements, which were quickly dismantled after the conquest in the region, a situation much different from the *limes*. Sanctuaries may also have had a role in the rural economy, as their cultural importance stems from and profits from the rural settlements. Indeed, they were supported by the local communities which found and maintain them, and they also provided permanent or temporary market places. Nevertheless, it is very difficult to distinguish clearly a sanctuary and a secondary agglomeration in the Somme, for the typical architecture of the *fanum* can be found in isolation or associated with numerous structures, be they commercial or dwellings. Without proper excavation of these settlements, it cannot be confirmed if in some cases the *fanum* is created by the secondary agglomeration/village/*vicus* or if the opposite applies. For this reason, they will not be distinguished but joined in a single variable.

The same approach which was employed for the *capita civitatum* is also applied here, with a variable created for the first half of the second century and another one for the first half of the fourth century. Additionally, a third version will first be employed, including all local foci, regardless of the lack of chronological data, which will be tested with all 822 rural sites. For the variable of the setting in the first half of the second century, the loss of sites is relatively small compared to the massive loss in rural settlements: 16 secondary agglomerations and sanctuaries were ignored among a total of 75 in the Somme department. The changes brought with the fourth century are the loss of 24 secondary agglomerations, but two sites are also created in the third century and are occupied during the fourth century. In total there are therefore 60 secondary agglomerations during the second century and 38 in the fourth.

The reclassification of the model is different from the one of the *capita*, as several hours of travel would not be as sought after in direction of these local foci than towards the main cities. The most predictive category should therefore be located half an hour cost at most from the local foci, the medium one up to one hour and the last one above one hour cost away.

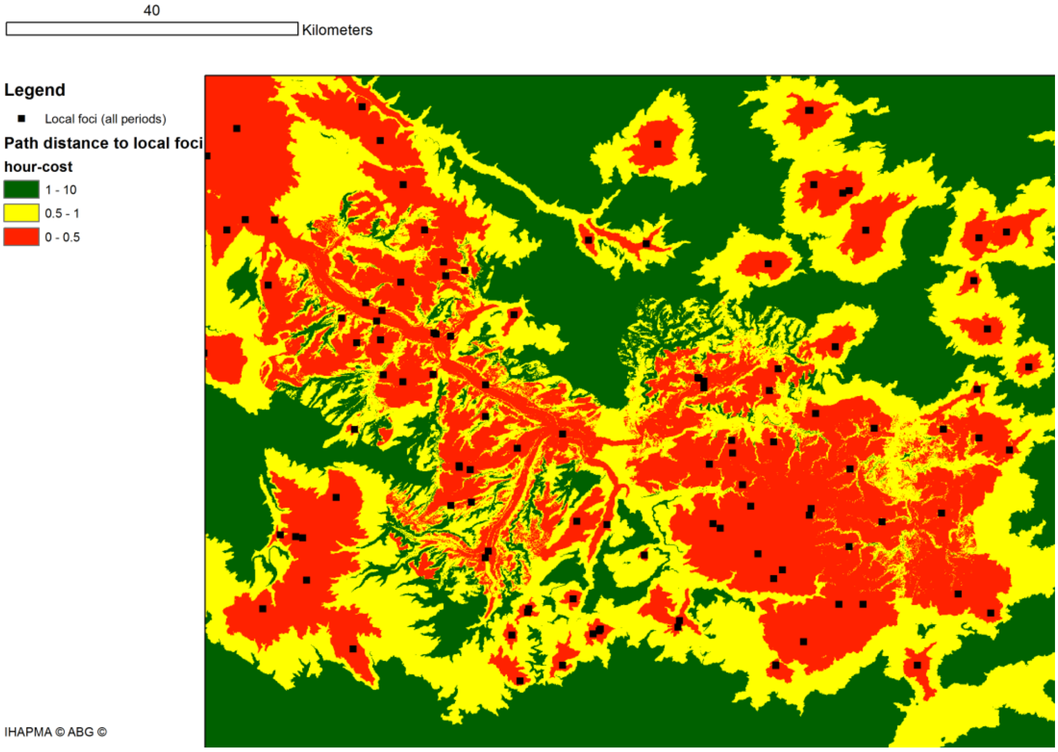


Figure 25: Bivariate model of path distance to local foci and the physical landscape (no specific period) , IHAPMA© ABG©

The results of the model on all sites give very good results everywhere but region four (fig. 25 and table 16). Indeed, in all first three regions each class is very well defined by heterogeneous predictive values in relatively balanced areas in terms of surface. Only the fourth region strays behind, as its low predictive class' surface is strongly diminished, which should be imputed to the path distance algorithm's behaviour with this very flat area.

Notwithstanding this specific region, the model's results conform particularly well to the theory and therefore imply a strong correlation between rural settlement location and accessibility from local foci, be they sanctuaries, *vici* or secondary agglomerations. This potential influence is felt at mid-distance especially, because the high predictive area is not improved much compared to the model of the physical landscape alone, the only gains being made in the second region, whereas the medium predictive class is optimised in all regions.

Path distance from local foci	Predictive class	Observed sites	Expected	P_o	Area ratio P_a	Kvamme's gain	Indicative value	Relative gain	Chi-square test
Region 1									
	1-10 h	6	23.00	0.03	0.11	-2.83 (-2.09)	0.26 (-0.31)	-0.08 (-0.03)	0.00000
	0.5-1 h	49	65.46	0.24	0.31	-0.34 (+0.20)	0.75 (+0.10)	-0.08 (+0.04)	
	0-0.5 h	153	119.54	0.74	0.57	0.22 (-0.01)	1.28 (-0.02)	0.16 (-0.01)	
Region 2									
	1-10 h	7	27.06	0.04	0.16	-2.87 (-2.63)	0.26 (-0.55)	-0.12 (-0.10)	0.00001
	0.5-1 h	62	65.63	0.36	0.38	-0.06 (+0.25)	0.94 (+0.18)	-0.02 (+0.06)	
	0-0.5 h	102	78.31	0.60	0.46	0.23 (+0.07)	1.30 (+0.11)	0.14 (+0.04)	
Region 3									
	1-10 h	12	30.88	0.06	0.15	-1.57 (-0.40)	0.39 (-0.07)	-0.09 (-0.04)	0.00000
	0.5-1 h	56	74.09	0.27	0.36	-0.32 (+0.61)	0.76 (+0.24)	-0.09 (+0.08)	
	0-0.5 h	136	99.03	0.67	0.49	0.27 (-0.02)	1.37 (-0.04)	0.18 (-0.05)	
Region 4									
	1-10 h	2	1.31	0.01	0.01	0.35 (+0.70)	1.53 (+0.79)	0.00 (+0.02)	0.68752
	0.5-1 h	61	57.09	0.26	0.24	0.06 (+0.96)	1.07 (+0.54)	0.02 (+0.14)	
	0-0.5 h	176	180.60	0.74	0.76	-0.03 (-0.20)	0.97 (-0.24)	-0.02 (-0.16)	

Table 16: Results of the bivariate model of the path distance to local foci and the physical landscape (no specific period), ABG© IHAPMA©

The same trend visible with all sites is also clearly defined in the first half of the second century (fig. 26 and table 17), but with an even worse situation in the fourth region, where the complete inversion of values in predictive classes are difficult to explain other than by the absence of some local foci, removed during the chronological filtering.

In the fourth century (fig.27 and table 18), the situation is surprisingly the same, with even stronger heterogeneity between classes. Even if the number of sites is extremely low, the near-absence of sites in the low predictive class – although it expanded in this period – may indicate not only a strong influence between rural settlement location and local foci, but maybe even a dependence, apart again in the fourth region. Did sites survive longer during the Late Roman period thanks to the proximity of local foci? If the site corpus for the period were larger, an answer to this question might possibly be given.

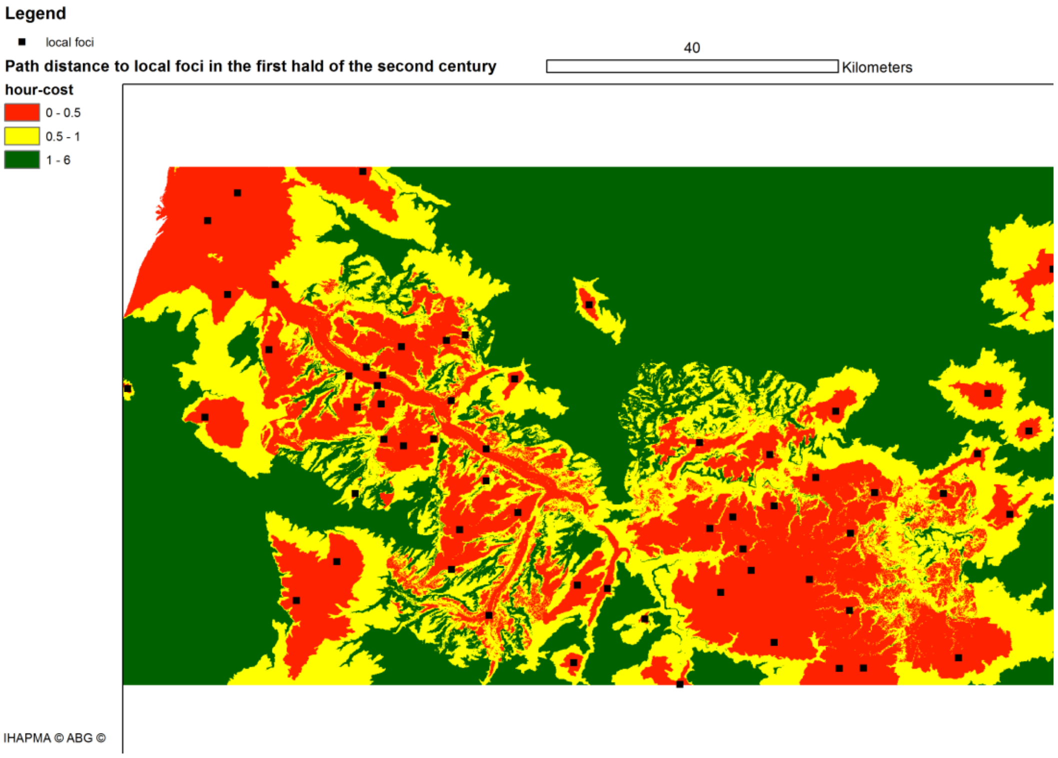


Figure 26: Bivariate model of the path distance to local foci and the physical landscape during the first half of the second century, IHAPMA© ABG©

Path distance from local foci (2nd century)	Predictive class	Observed sites	Expected	P_o	Area ratio P_a	Kvamme's gain	Indicative value	Relative gain	Chi-square test
Region 1									
	1	1	10.88	0.01	0.14	-9.88	0.09	-0.13	0.00004
	2	18	26.59	0.23	0.35	-0.48	0.68	-0.11	
	3	58	39.53	0.75	0.51	0.32	1.47	0.24	
Region 2									
	1	3	15.24	0.04	0.19	-4.08	0.20	-0.15	0.00171
	2	37	33.55	0.47	0.42	0.09	1.10	0.04	
	3	39	30.21	0.49	0.38	0.23	1.29	0.11	
Region 3									
	1	6	14.69	0.08	0.19	-1.45	0.41	-0.11	0.03341
	2	33	31.14	0.43	0.41	0.06	1.06	0.02	
	3	37	30.17	0.49	0.40	0.18	1.23	0.09	
Region 4									
	1	7	1.80	0.11	0.03	0.74	3.88	0.08	0.00003
	2	29	21.90	0.45	0.34	0.24	1.32	0.11	
	3	28	40.30	0.44	0.63	-0.44	0.69	-0.19	

Table 17: Results of the bivariate model of the path distance to local foci and the physical landscape (first half of the second century), ABG© IHAPMA©

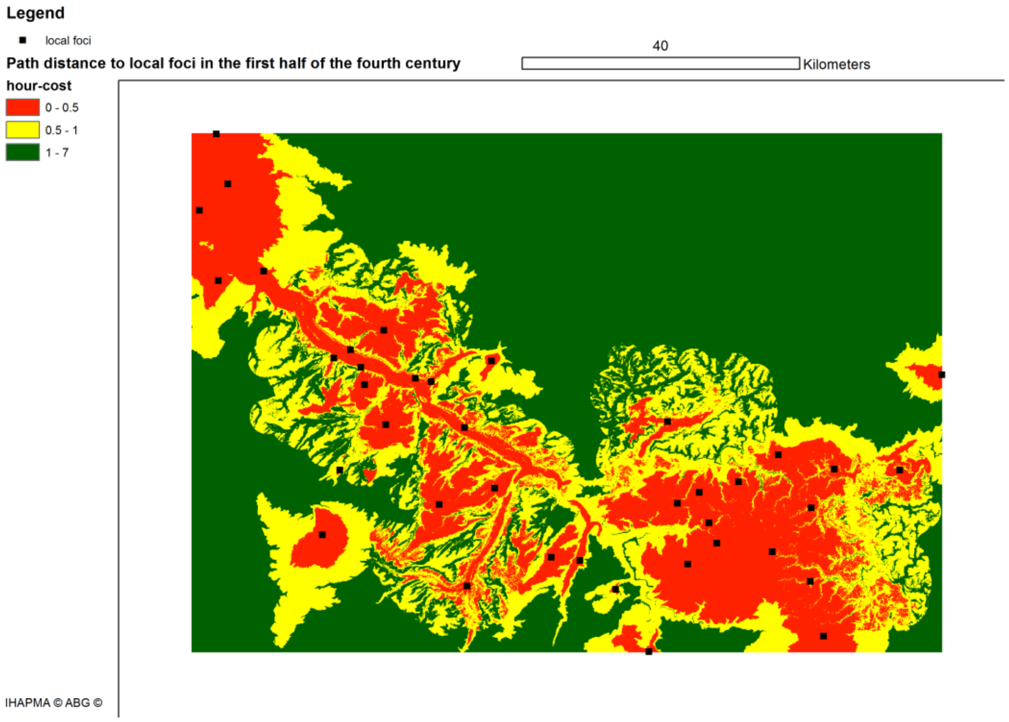


Figure 27: Bivariate model of the path distance to local foci and the physical landscape during the first half of the fourth century, IHAPMA© ABG©

Path distance to local foci in the from local foci (4th century)	Predictive class	Observed sites	Expected	P_o	Area ratio P_a	Kvamme's gain	Indicative value	Relative gain	Chi-square test
Region 1									
	1	0	4.27	0.00	0.20	0.00	0.00	-0.20	0.02234
	2	8	8.81	0.38	0.42	-0.10	0.91	-0.04	
	3	13	7.92	0.62	0.38	0.39	1.64	0.24	
Region 2									
	1	0	6.51	0.00	0.20	0.00	0.00	-0.20	0.00695
	2	14	14.02	0.42	0.42	0.00	1.00	0.00	
	3	19	12.46	0.58	0.38	0.34	1.52	0.20	
Region 3									
	1	1	7.65	0.03	0.23	-6.65	0.13	-0.20	0.00128
	2	13	15.27	0.38	0.45	-0.17	0.85	-0.07	
	3	20	11.07	0.59	0.33	0.45	1.81	0.26	
Region 4									
	1	3	1.96	0.11	0.07	0.35	1.53	0.04	0.51206
	2	13	11.33	0.48	0.42	0.13	1.15	0.06	
	3	11	13.71	0.41	0.51	-0.25	0.80	-0.10	

Table 18: Results of the bivariate model of the path distance to local foci and the physical landscape (first half of the fourth century), ABG© IHAPMA©

3.3.4 Multivariate model of distance to archaeological features

The last three models will now be combined in a weighted multivariate model, together with the model of the physical landscape. The theory is that the physical landscape must have a slightly lower weight than all socio-economic variables together and in the latter, the most influence will be taken by the local foci, then the cities and lastly the roads. As there are four variables and the result of the combination would have to be divided by four, the total of weights must also be four if the classification is to be between 1 and 3. Here is the equation for this model:

(Physical landscape X 1.6 + Distance to local foci X 1 + Distance to cities X 0.8 + Distance to roads X 0.6) / 4 = best probability of site location in flat landscape, near local foci, accessible from cities and – maybe but not restrictively – close to the main roads.

Here are the results of the model, displaying global parameters (no specific period), then the first half of the second century and finally the first half of the fourth century:

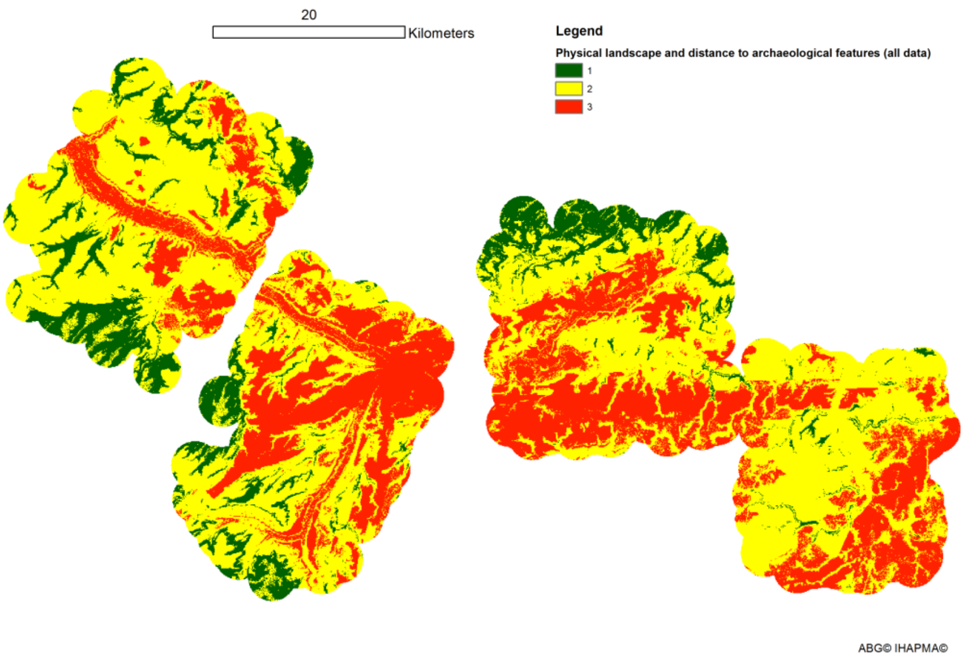


Figure 28: Multivariate weighted model of the distance to archaeological features and the physical landscape (no specific period), IHAPMA© ABG©

The model’s performance using all data is contrast-full (fig.28 and table 19): classes generally differ widely in predictive values, apart from the fourth region which yet again displays a near homogeneity of site location according to the very close relative gain and indicative values. This directly impacts the Chi-square test, and therefore displays a very high randomness probability. In other regions, the low predictive class, although reduced in surface, has extremely low predictive values; just has the high predictive class gets relatively high values. The Chi-square test is excellent in these three regions and the inversion of the medium and high predictive values in the first model does not contradict with the theory, as both display favourable locations, in terms of physical environment and distance to archaeological focal points. The most positive aspect of the model regarding the theory is the strong deviation of all values from the model only based on the physical landscape. Therefore, despite having been weighted more fairly than the physical landscape variable, the socio-economic variables have a very strong influence on the model's behaviour and quite possibly on site location.

Physical landscape and distance to archaeological features (all data)	**Predictive class**	**Observed sites**	**Expected**	**P_o**	**Area ratio P_a**	**Kvamme's gain**	**Indicative value**	**Relative gain**	**Chi-square test**
Region 1									
	1	9	36.30	0.04	0.17	-3.03 (-2.29)	0.25 (-0.32)	-0.13 (-0.28)	0.00000
	2	166	135.04	0.80	0.65	0.19 (+0.73)	1.23 (+0.58)	0.15 (+0.27)	
	3	33	36.67	0.16	0.18	-0.11 (-0.34)	0.90 (-0.40)	-0.02 (-0.19)	
Region 2									
	1	2	18.17	0.01	0.11	-8.09 (-7.85)	0.11 (-0.70)	-0.09 (-0.07)	0.00000
	2	71	84.20	0.42	0.49	-0.19 (+0.12)	0.84 (+0.08)	-0.08 (0.00)	
	3	98	68.62	0.57	0.40	0.30 (+0.14)	1.43 (+0.24)	0.17 (+0.07)	
Region 3									
	1	12	28.36	0.06	0.14	-1.36 (-0.19)	0.42 (-0.04)	-0.08 (-0.03)	0.00000
	2	72	96.66	0.35	0.47	-0.34 (+0.59)	0.74 (+0.22)	-0.12 (+0.05)	
	3	120	78.98	0.59	0.39	0.34 (+0.05)	1.52 (+0.11)	0.20 (-0.03)	
Region 4									
	1	4	5.35	0.02	0.02	-0.34 (+0.01)	0.75 (+0.01)	-0.01 (+0.01)	0.83049
	2	151	151.28	0.63	0.63	0.00 (+0.90)	1.00 (+0.47)	0.00 (+0.12)	
	3	84	82.37	0.35	0.34	0.02 (-0.15)	1.02 (-0.19)	0.01 (-0.13)	

Table 19: Results of the multivariate weighted model of the distance to archaeological features and the physical landscape (no specific period), ABG© IHAPMA©

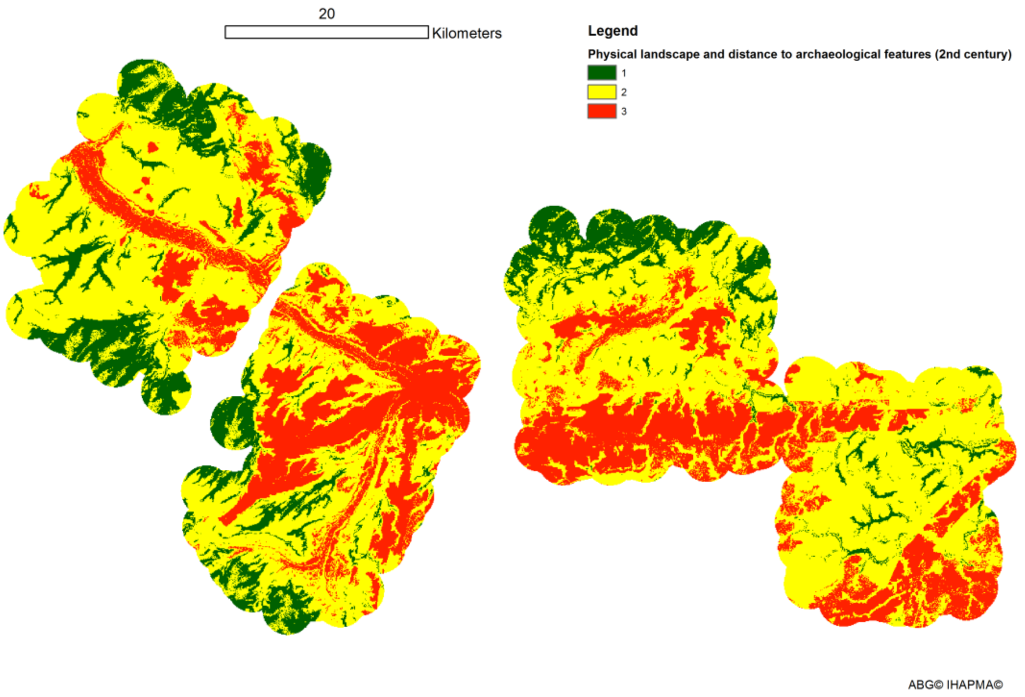


Figure 29: Multivariate weighted model of the distance to archaeological features and the physical landscape (first half of the second century), IHAPMA© ABG©

Physical landscape and distance from archaeological features (2nd century)	Predictive class	Observed sites	Expected	P_o	Area ratio P_a	Kvamme's gain	Indicative value	Relative gain	Chi-square test
Region 1									
	1	4	16.66	0.05	0.22	-3.17	0.24	-0.16	0.00204
	2	59	47.88	0.77	0.62	0.19	1.23	0.14	
	3	14	12.46	0.18	0.16	0.11	1.12	0.02	
Region 2									
	1	3	9.51	0.04	0.12	-2.17	0.32	-0.08	0.00057
	2	31	40.18	0.39	0.51	-0.30	0.77	-0.12	
	3	45	29.32	0.57	0.37	0.35	1.53	0.20	
Region 3									
	1	2	11.11	0.03	0.15	-4.56	0.18	-0.12	0.00311
	2	40	40.78	0.53	0.54	-0.02	0.98	-0.01	
	3	34	24.11	0.45	0.32	0.29	1.41	0.13	
Region 4									
	1	8	3.24	0.13	0.05	0.59	2.47	0.07	0.02510
	2	41	43.88	0.64	0.69	-0.07	0.93	-0.04	
	3	15	16.88	0.23	0.26	-0.13	0.89	-0.03	

Table 20: Results of the multivariate weighted model of the distance to archaeological features and the physical landscape (first half of the second century), ABG© IHAPMA©

In the second century (fig.29 and table 20), the model's results are similar to the preceding one, apart from the complete inversion of predictive values in the fourth region, which is in part due to the surface of the low prediction area being only 5% of the region's area; but it may also indicate that site location is not limited by the input variables at all.

Finally, the fourth century's trends – still very debatable considering the small amount of sites for this period – still agree with the global trends (fig.30 and table 21) and if the low predictive values in the low prediction class are representative of the area's trends in the fourth century, then site location, or rather site survival during this period is heavily correlated with attractive parameters, even in the fourth region which now follows this model despite a poor Chi-square result. Were rural settlements more prone to abandonment in areas remote from cities, roads or local foci? It could be the case, but as previously mentioned, much more data is needed to confirm this assumption.

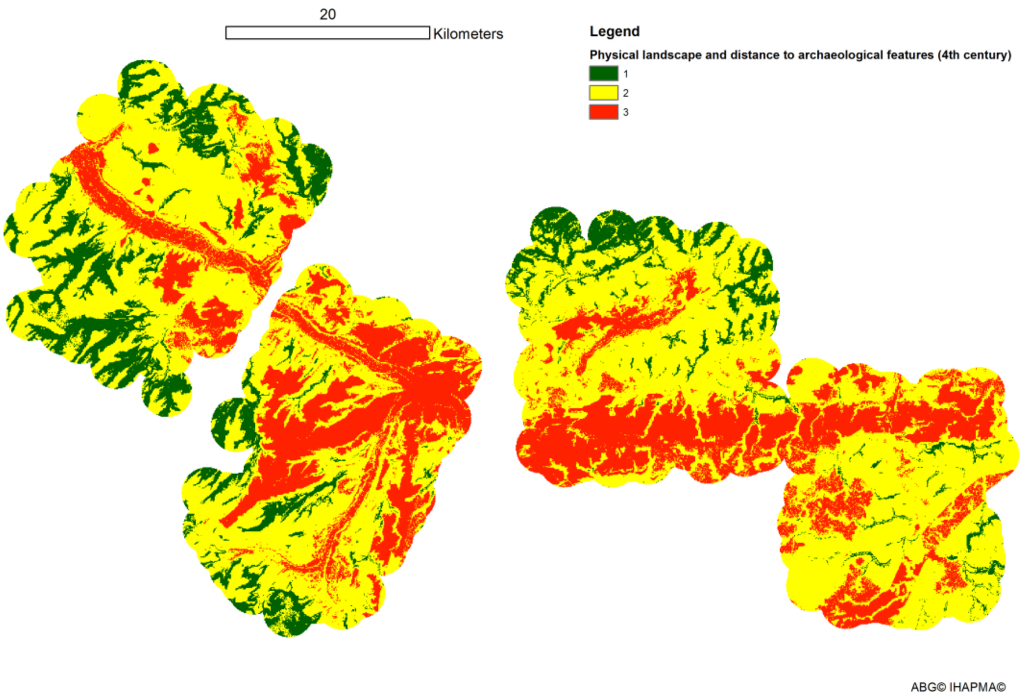


Figure 30: Multivariate weighted model of the distance to archaeological features and the physical landscape (first half of the fourth century), IHAPMA© ABG©

Physical landscape and distance to archaeological features (4th century)	**Predictive class**	**Observed sites**	**Expected**	P_o	**Area ratio** P_a	**Kvamme's gain**	**Indicative value**	**Relative gain**	**Chi-square test**
Region 1									
	1	2	4.70	0.10	0.22	-1.35	0.43	-0.13	0.27935
	2	14	13.04	0.67	0.62	0.07	1.07	0.05	
	3	5	3.26	0.24	0.16	0.35	1.53	0.08	
Region 2									
	1	0	3.28	0.00	0.10	0.00	0.00	-0.10	0.00345
	2	12	17.51	0.36	0.53	-0.46	0.69	-0.17	
	3	21	12.21	0.64	0.37	0.42	1.72	0.27	
Region 3									
	1	0	4.36	0.00	0.13	0.00	0.00	-0.13	0.00007
	2	13	19.72	0.38	0.58	-0.52	0.66	-0.20	
	3	21	9.91	0.62	0.29	0.53	2.12	0.33	
Region 4									
	1	0	0.81	0.00	0.03	0.00	0.00	-0.03	0.64903
	2	18	17.87	0.67	0.66	0.01	1.01	0.00	
	3	9	8.32	0.33	0.31	0.08	1.08	0.03	

Table 21: Results of the multivariate weighted model of the distance to archaeological features and the physical landscape (first half of the fourth century), ABG© IHAPMA©

3.3.5 Conclusions on the influence of archaeological features

It can be concluded that the three socio-economic variables used in combination with the model of the physical landscape have a very important correlation with site location, in all periods. To determine if this correlation is representative of actual decision making on the part of Roman landowners is very conjectural. Though, any deterministic process should be excluded from the interpretation of the results, especially in the fourth micro-region, whose fertile flat plateaus and very good setting regarding all parameters allows the sites to be located very randomly and therefore homogeneously in the area. In the other areas, where the local physical landscape is more contrasted, site location is greatly correlated with favourable places and the neighbourhood of economic focal points.

It may also be surmised from the results of the preceding models that rural settlement location may be more influenced by the presence or absence of sanctuaries, *vici* and secondary agglomerations, than by the main cities. The correlation with very close roads is a lesser one, but its influence may still hold, which may be more telling in the spatial arrangement along their path rather than on the statistical scale.

In a general way, the physical landscape model that was created, if it is slightly correlated with site location probabilities, compares less to socio-economic factors which impart much wider variations and heterogeneities. This conclusion does not refute all past hypothesis made on site location in the area, even though they were produced through monographic studies or empirical claims. Potential divergences between types of rural settlements regarding the most pertinent variables and models will now be inquired.

3.4 Typological divergences

Until now, rural settlements have been analysed from a very globalised perspective, making no distinction between different typologies of sites in the models. Site location may in fact be different for different classes of sites: large villas may develop in the best possible locations, while post-built settlements may be located in less favourable areas. This is yet to be proved, or at least correlated with the results of the models. The potential divergences between types of sites in the four micro-regions will be explained through the ratios of observed sites per relevant model and per types of sites, including all sites (already presented in the preceding chapter), post-built settlements, stone-built settlements and villas. These tables concern slopes (table 22), DEV landforms (table 23), path distance to water (table 24), the multivariate model of physical landscape (table 25), Euclidian distance to Roman roads (table 26), path distance to *capita civitatum* (table 27), path distance to local foci (table 28) and the multivariate model of physical landscape combined with distance to archaeological features (table 29).

First of all, as stone-built sites represent the vast majority of rural settlement in the micro-regions (729 sites or 89%), they of course constitute the most representative element of this dataset and all their ratios are near-identical to the ones showcased in Chapter 3. Their site location characteristics are therefore the same ones underlined in the preceding chapter and will not be described again. The following results describe the variations in distribution of post-built settlements and villas from the trends of stone-built settlements.

For post-built rural settlements and villas, although the dataset is extremely limited – 61 sites and 32 sites respectively – important divergences or inversely unexpected concordance with the trends of stone-built settlements can be noted. Nonetheless a very critical view of these ratios must be kept, as they could radically change through the inclusion of a handful of newly discovered sites in either class.

Slope classes	All sites	Post-built	Stone-built	Villas
Region 1				
6% - 82%	0.07	0.20	0.07	0.00
2% - 6%	0.09	0.05	0.10	0.14
0% - 2%	0.84	0.75	0.83	0.86
Region 2				
6% - 82%	0.06	0.12	0.06	0.07
2% - 6%	0.09	0.06	0.10	0.00
0% - 2%	0.85	0.82	0.84	0.93
Region 3				
6% - 82%	0.05	0.00	0.05	0.08
2% - 6%	0.09	0.15	0.07	0.08
0% - 2%	0.86	0.85	0.88	0.84
Region 4				
6% - 82%	0.02	0.00	0.02	0.00
2% - 6%	0.03	0.00	0.04	0.00
0% - 2%	0.95	1.00	0.94	1.00

Table 22: Proportions of observed sites per slope class and per type of settlement, ABG© IHAPMA©

DEV Landforms	All sites	Post-built	Stone-built	Villas
Region 1				
Deep valley	0.03	0.00	0.03	0.00
Hilltops/ lower slopes	0.16	0.15	0.14	0.57
Flat/mid/ upper slopes	0.81	0.85	0.83	0.43
Region 2				
Deep valley	0.06	0.18	0.06	0.00
Hilltops/ lower slopes	0.25	0.53	0.22	0.24
Flat/mid/ upper slopes	0.70	0.29	0.72	0.76
Region 3				
Deep valley	0.02	0.06	0.02	0.00
Hilltops/ lower slopes	0.13	0.31	0.11	0.08
Flat/mid/ upper slopes	0.85	0.62	0.87	0.92
Region 4				
Deep valley	0.06	0.00	0.06	0.00
Hilltops/ lower slopes	0.12	0.20	0.12	0.00
Flat/mid/ upper slopes	0.82	0.80	0.82	1.00

Table 23: Proportions of observed sites per DEV landform class and per type of settlement, ABG© IHAPMA©

Path-distance to rivers (hours)	All sites	Post-built	Stone-built	Villas
Region 1				
0 - 0.0166	0.03	0.00	0.03	0.14
0.0166 - 0.5 / >1	0.27	0.26	0.25	0.71
0.5 - 1	0.70	0.74	0.71	0.14
Region 2				
0 - 0.0166	0.06	0.07	0.05	0.20
0.0166 - 0.5 / >1	0.44	0.87	0.41	0.30
0.5 - 1	0.49	0.07	0.53	0.50
Region 3				
0 - 0.0166	0.01	0.07	0.01	0.00
0.0166 - 0.5 / >1	0.36	0.40	0.36	0.33
0.5 - 1	0.63	0.53	0.63	0.67
Region 4				
0 - 0.0166	0.01	0.00	0.01	0.00
0.0166 - 0.5 / >1	0.94	1.00	0.94	0.86
0.5 - 1	0.05	0.00	0.05	0.14

Table 24: Proportions of observed sites per class of path-distance to rivers and per type of settlement, ABG© IHAPMA©

Physical Landscape	All sites	Post-built	Stone-built	Villas
Region 1				
1	0.07	0.00	0.07	0.14
2	0.21	0.37	0.19	0.43
3	0.72	0.63	0.74	0.43
Region 2				
1	0.09	0.27	0.08	0.00
2	0.26	0.40	0.24	0.30
3	0.65	0.33	0.68	0.70
Region 3				
1	0.04	0.13	0.04	0.00
2	0.19	0.20	0.19	0.17
3	0.77	0.67	0.77	0.83
Region 4				
1	0.06	0.00	0.06	0.00
2	0.13	0.14	0.13	0.00
3	0.81	0.86	0.80	1.00

Table 25: Proportions of observed sites per class of the multivariate model of the physical landscape and per type of settlement, ABG© IHAPMA©

Euclidian distance to Roman roads	**All sites**	**Post-built**	**Stone-built**	**Villas**
Region 1				
2000-97870 m	0.26	0.37	0.24	0.57
1000-2000 m	0.69	0.47	0.72	0.43
0-1000 m	0.05	0.16	0.04	0.00
Region 2				
2000-97870 m	0.23	0.33	0.23	0.20
1000-2000 m	0.39	0.40	0.39	0.30
0-1000 m	0.38	0.27	0.38	0.50
Region 3				
2000-97870 m	0.13	0.20	0.14	0.00
1000-2000 m	0.46	0.33	0.47	0.33
0-1000 m	0.41	0.47	0.39	0.67
Region 4				
2000-97870 m	0.10	0.00	0.10	0.00
1000-2000 m	0.65	0.71	0.65	0.86
0-1000 m	0.25	0.29	0.25	0.14

Table 26: Proportions of observed sites per class of Euclidian distance to Roman roads and per type of settlement, ABG© IHAPMA©

Path distance to *capita civitatum*	**All sites**	**Post-built**	**Stone-built**	**Villas**
Region 1				
2.25-10 h	0.10	0.10	0.10	0.00
1.75-2.25 h	0.58	0.58	0.59	0.14
0-1.75h	0.32	0.32	0.30	0.86
Region 2				
2.25-10 h	0.02	0.00	0.02	0.00
1.75-2.25 h	0.26	0.47	0.26	0.00
0-1.75h	0.72	0.53	0.72	1.00
Region 3				
2.25-10 h	0.02	0.00	0.02	0.00
1.75-2.25 h	0.24	0.53	0.21	0.25
0-1.75h	0.75	0.47	0.77	0.75
Region 4				
2.25-10 h	0.06	0.00	0.06	0.00
1.75-2.25 h	0.66	0.71	0.65	0.86
0-1.75h	0.28	0.29	0.29	0.14

Table 27: Proportions of observed sites per class of path distance to *capita civitatum* and per type of settlement, ABG© IHAPMA©

Path distance to local foci	All sites	Post-built	Stone-built	Villas
Region 1				
1-10 h	0.03	0.00	0.03	0.00
0.5-1 h	0.24	0.21	0.24	0.29
0-0.5 h	0.74	0.79	0.73	0.71
Region 2				
1-10 h	0.04	0.07	0.04	0.00
0.5-1 h	0.36	0.60	0.35	0.20
0-0.5 h	0.60	0.33	0.61	0.80
Region 3				
1-10 h	0.06	0.07	0.06	0.00
0.5-1 h	0.27	0.33	0.25	0.50
0-0.5 h	0.67	0.60	0.68	0.50
Region 4				
1-10 h	0.01	0.00	0.01	0.00
0.5-1 h	0.26	0.29	0.24	0.29
0-0.5 h	0.74	0.71	0.75	0.71

Table 28: Proportions of observed sites per class of path distance to local foci and per type of settlement, ABG© IHAPMA©

Physical landscape and distance to archaeological features	All sites	Post-built	Stone-built	Villas
Region 1				
1	0.04	0.00	0.05	0.00
2	0.80	0.79	0.80	0.71
3	0.16	0.21	0.15	0.29
Region 2				
1	0.01	0.00	0.01	0.00
2	0.42	0.53	0.42	0.10
3	0.57	0.47	0.56	0.90
Region 3				
1	0.06	0.13	0.06	0.00
2	0.35	0.27	0.35	0.50
3	0.59	0.60	0.59	0.50
Region 4				
1	0.02	0.00	0.01	0.00
2	0.63	0.57	0.63	0.86
3	0.35	0.43	0.36	0.14

Table 29: Proportions of observed sites per class of the multivariate model of the physical landscape and the distance to archaeological features and per type of settlement, ABG© IHAPMA©

Post-built settlements display globally similar location preferences to stone-built settlements in terms of variables in the physical landscape, although it may be noted that in regions 1 and 2, they are slightly more represented in areas of important slopes and in the lower slopes of valleys (tables 22 and 23), the latter constituting the main class for post-built sites in region 2. In this second region different hydrological preferences can also be seen, with a strong presence near river courses and very far from them (table 24). More differences are outlined when considering the distance to archaeological features: post-built settlements of regions 1, 2 and 3 are more remotely located in regards to Roman roads (table 26); the same goes with region 2 and 3's post-built settlements which are further from *capita civitatum* or region 2's to local foci (tables 27 and 28). Thus, if globally similar to the results of stone-built settlements, post-built sites of regions 2 and 3 may still indicate a slightly more remote site location – both in terms of physical and socio-economic landscape – as well as a faintly higher preference for the lower slopes of valleys (tables 25 and 29). Region 4 may display a quaint opposing trend and favour flatter areas and higher accessibility to archaeological features.

The villas, the least represented class in quantitative terms, but certainly the one with the best detection rate during aerial and field surveys, should be scrutinised further. There is not a single villa located in deep valley bottoms but in region 2 they are often located in lower slopes rather than very flat areas, which is also the case for the other three regions (table 23). In a fashion contradictory to the presupposed need for large and dry open spaces, villas may show a slightly more frequent site location in the immediate vicinity of rivers (23), in regions 1 and 2. Although the villas of region 4 are exclusively located in the plateaus and region 2's are more often in the lower slopes of valleys, the other two regions indicate a strong analogy with the location of stone-built settlements in the physical landscape (table 25). Just as for post-built sites, stronger divergences are evident when examining socio-economic variables: if the first region (whose road segment is very limited) displays a preferential location far from Roman roads, the other three regions' villas are close to them (table 26). The beneficial influence of a higher accessibility to *capita civitatum* and local foci is very pronounced in all regions (tables 27 and 28). As a result, the overall accessibility to archaeological points of interest is adamant for the site location of the Somme’s villas, although a moderate accessibility is preferred in most regions but the second, where the dense Road network and the presence of Amiens may skew the results towards a superior proximity.

Now the implications of the models’ results shall be discussed further, as well as the benefits and issues regarding the methodology which was employed during this work.

Chapter 4: Discussion

The results of the models showed that important correlations between rural settlement location and specific variables or combinations of variables can be made. The following chapter will discuss in greater detail the outcomes of the previous models regarding the study of settlement patterns in Northern Gaul, as well as the methodological implications of this work and the issues left to be solved.

4.1. Rural settlement patterns

4.1.1 Post-built settlements, an invisible majority?

The 61 post-built settlements of this dataset seem like only a handful compared to stone-built settlements, but this may very well be the reflection of multiple biases on the detection and observation of such sites. Firstly, this classification only represents what could be the 'optimal' development of each settlement (Durand-Dastès *et al.* 1998, 161–162). Secondly, as most data originates from aerial surveys, the most visible phases of construction are generally stone buildings and enclosures and therefore take precedence in the descriptions.

Many post-built rural settlements remain unknown, for the earliest phases of stone-built settlements and villas are often constituted by post-built 'native farms' with complex organisations, sometimes lasting many generations, such as Behen-Huchenneville in the Somme or in Conchil-le-Temple and Hoogeloon further north (Bayard *et al.* 2014, 40–45; Lemaire *et al.* 1996; Roymans *et al.* 2015; Reddé 2016, 31–32). Late Roman phases also frequently include the complete dismantlement of masonries, replaced by perishable structures. Indeed, stone-built settlements such as the one in Pont-de-Metz include Late Iron Age and Early Roman rectangular enclosures with small post-built buildings as well as Late Roman wooden constructions (Lemaire 1994; Revert 2016, 94–95). In the absence of excavation, such structures would barely have been noticed. In morphological terms, post-built settlements later than the first half of the 1st century AD also often display very vast built areas (more than 4 ha), with symmetrical and orderly organisations, in fact very reminiscent of villa settlements (Pichon 2009, 325). Post-built settlements of the studied period directly inherit the morphology of many Late Iron Age ones with aligned courtyards, which are in fact considered by many as the ancestors of the Gallo-Roman style of villa, which differs strongly from the Mediterranean models (Fichtl 2009).

Roger Agache clearly understood that site location analysis of post-built settlements – not exclusively of the Roman period – is a very difficult endeavour. He noticed that those which he detected were always located on poor chalky slopes, but that this actually does not represent their most common location, for their detection is mainly possible on these chalky soils, just as the loess surfaces of the plateaus are favourable for the detection of stone buildings (Agache 1976, 122–123). Strong doubts should be emitted towards the results of the analyses regarding these kinds of settlements, as they may indicate a detection potential rather than a representative occurrence.

Nevertheless, with recent rescue excavations, more post-built settlements can be found, notably in non-chalky soils. Indeed, the large-scale excavations following the construction of the Canal-Seine-Nord uncovered mainly post-built sites (5 of the 9 Roman settlements), even if the landscape is one of silt and alluvial deposits (see fig. 9 above). The same goes with the excavations preceding the construction of the aero-industrial platform of Méaulte, in the third micro-region, where the four discovered sites were post-built and very similar to Late Iron Age examples (Duvette 2008; Ouzoulias 2009, 153). In the Selle valley, very recent excavations also indicate a large proportion of post-built settlements, as well as wooden buildings of different phases, quantitatively dominant in all sites, including the villa of Pont-de-Metz (Blondiau 2014, 7–11). These settlements do not indicate a lag in the Roman influence and it is still very speculative to associate their architecture with a low social status for their owners, as was carried out by Blaise Pichon (2009, 237). Pierre Ouzoulias depicts them as representing peasant families, economically autonomous and very possibly owning the land they live in (Ouzoulias 2009, 154).

Despite these modest architectural modalities, all sites benefit from very favourable locational parameters: ‘on the border of the plateau, they dominate the Selle river (...) navigable by means of barges or small boats, ... They are inserted in a network of past and contemporary sites (...), the plateau is fertile’ (Blondiau 2014, 13). In future research, the interrelation of these well-excavated examples with field and aerial survey data could provide a possible method for sketching out a more accurate representation of post-built settlements. Nonetheless, theories about post-built settlements being located in very marginal areas of the Somme can be discarded, although as shown by the models' results, they can indeed exist in places remote from the main economic foci of the landscape.

4.1.2 A dense pattern of stone-built settlements

Not only were non-villa stone-built settlements more numerous than any other type of settlements, but they also seem to be responsible for the vast majority of the rural productions. In the vicinity or Reims, an eloquent settlement pattern has now been evidenced multiple times: although this city was the most important one in *Gallia Belgica Prima*, nearly no villa could be found in a 15 km radius around the *caput civitatis* and very few further (Reddé 2016, 18; Ouzoulias 2011, 483). This 'paradoxical' pattern goes against the presence of many nobles in the city and traditional conceptions of the Roman rural world: a non-villa landscape may have indeed existed in heavily Romanised areas. Despite this largely non-villa landscape of stone-built settlements, the Reims region was sufficiently efficient to be noted by Pliny as one of the main producers of free-threshing wheat, which clearly indicates that the organisation of production and the soils were highly suitable for this typically Mediterranean style of extensive production (Ouzoulias 2011, 484).

The relative lack of literature regarding this class of settlements can be explained by its very wide diversity of shapes and chronological evolutions. Just as for post-built settlements, the strong emphasis given in the past and the present to villa settlements is also detrimental to a proper study of stone-built settlements. In this study, this class of settlements was discriminated by mentioning the use of stone and the lack of evidence for them being 'true' villas (a very vast settlement built of stone, with a residential area containing many rooms and a large number of agricultural buildings). It is true that the distinction between many stone-built settlements and villas is still very troublesome and recent authors still regard the model of the villa as the dominating factor in the Roman rural landscape (Pichon 2009, 237). Many rural settlements also display a very mixed architecture, with a few stone buildings cohabitating with a multitude of post-built constructions, even in villas. The only certainty is that architecture changed a lot during the studied period; stone was used mostly in foundations, while it is still uncertain if stone was used in the elevation of walls.

Materials used in foundation are not always representative of a higher grade of architecture. When archaeological ground levels are well preserved by field ploughing, settlements which at first sight could seem very modest, with small limestone foundations and wattle and daub elevation, could in fact receive painted plasters, for example in Dury or the earliest phases in the villa of Frémontiers, which also contain ground mosaics (Quérel *et al.* 2000; Vermeersch 1981, 150). Future research might shed light on such architectural developments in 'modest' stone-built settlements.

Van Ossel and Ouzoulias' research in the Parisian Plain, where the physical environment is very similar to the Somme's, shows the same demographic growth around the change of era (BC to AD), with slightly more continuity into the Late Roman period (Ouzoulias *et al.* 2009, 113). Departing from the quantitative approaches of the Archaeomedes program, their qualitative inventory of 20 rural settlements indicates that the frontier between the mainly post-built and partially stone-built settlements and the villas is extremely thin. All of these settlements indeed possess hints of permanent settlement: cellars, wells, abundant artefacts and a diversified *instrumentum*: an assemblage of tools and implements for the economic and domestic life. The same can be said for samples of settlements gathered from large excavations in the Oise, Marne and North departments, where small settlements are also the dominant or sometimes the exclusive mode of exploitation in the area (Ouzoulias *et al.* 2009, 116–117). Smaller settlements should therefore not be discarded as 'annexes' of the larger ones, as was done in the Archaeomedes program (Van der Leeuw *et al.* 2003, 229).

Overall, this very large and diverse category of sites, between villa-inspired courtyards with small stone-built residential buildings and small disorderly ensembles of small buildings cohabitating with post-built ones, would need much more work so as to be reclassified in a coherent and justified manner (Leveau 2009), which was not the purpose of this thesis. Though, it can already be indicated that these settlements are generally very well connected to the main economic focal points of the Somme and could be located in most areas, with a preference for the flatter areas of the landscape. Wightman was adamant that the rural and urban worlds not be antagonistic entities, but rather interconnected ones which profit from one another and whose development is concomitant (Wightman 1975, 623–624). This study generally concurs with that statement, as socio-economic foci were most certainly attractive factors for site location, especially villas. The same could be said in Eastern France, where Nüsslein's study shows that farms benefitted from the proximity of cities, although being closer than 3-4km did not improve site location (Nüsslein 2016b, 275); in this case the physical environment's variables also display important correlations with site location, especially in the valleys of these sometimes very rugged landscapes (Nüsslein 2016b, 324).

4.1.3 A 'villa landscape'?

Far from proving the enthusiastic affirmation of Roger Agache and Marcel Le Glay that the villa was the 'normal, or even exclusive way of using the land' (Agache 1978, 386; Le Glay 1975), this study rather tends to show that by taking a slightly more restrictive approach to the attribution of the term 'villa' to stone-built settlements, a much more nuanced picture of the Roman land-use can be created. In Agache's perspective, the villa is shown to be an isolated, hyper-specialised settlement, a 'centre of civilisation' which does not have many relations with the rest of the settlements, especially agglomerations (Agache 1978, 280). As noted by Pierre Ouzoulias, even by using the classification of rural settlements by Agache, in some of the most fertile areas of the Somme, such as West of Roye (slightly to the South of region four) 75 % of all settlements were not villas (Ouzoulias 2006, 148). Interestingly enough, region four similarly displays the lowest frequency of villas in the studied area (below 3%), despite its environmental homogeneity and relative flatness which does not impede extensive agricultural traditions. This could of course be due to a lack of data, but may also be representative of an actual trend in this region which is part of a slightly different physical landscape (the Santerre), as well as a different *civitas* (*Viromanduorum* rather than *Ambianorum*).

This lack of preponderance of villas in quantitative terms was also pointed out very early by Edith Wightman, who already claimed that the landscape was not exclusively exploited by and for villas (Wightman 1975, 650). That some other kinds of settlements may be 'annexes' of villas is only speculative and only very discrete and rare archaeological pieces of evidence – such as epigraphy – might prove or disprove this dependence link, which would only apply to the studied area. Unfortunately, Northern Gaul is well known for its relative lack of epigraphy, even in urban areas. More and more emphasis should be put towards more 'modest' types of settlements than villas, as the latter have already become the weakest in terms of quantitative representation. Though, as Pierre Ouzoulias and Peter Garnsey acknowledged, even if they are very few, villas still constitute a 'dominant exception' by their considerable economic development and their investment in luxury residential functions (Ouzoulias 2009, 154; Garnsey 1996, 78).

The presence of large villas, such as those in all four of the micro-regions that were studied, should be linked with extensive farming. This is not synonymous of higher productivity, but rather pertains to a different organisation of production and ownership. Indeed, free-threshing wheats may be thought of as an indication of very productive soils, as they require excellent overall conditions to be cropped, but they are in fact frequently found in very small settlements in the Parisian plain and in the Oise, thus giving evidence that a Romanised of production of surplus grain could be achieved by non-villa settlements (Reddé 2016, 35). In fact, it is too often surmised that an architectural development should be correlated with a 'use of land obeying to the principles of profitability defined by the Latin agronomists' (Leveau *et al.* 1999, 287–288). Though, in the Somme, this search for profitability is not clearly reflected in terms of site location, as the models' results have only shown a preference for areas very well connected to economic foci of the landscape.

Referring to van der Leuuw's classification of spatial systems (Van Der Leeuw 1981, 272–273), 'network' and 'solar systems' can safely be excluded as pertinent models for the Roman Somme, i.e. systems which respectively have no hierarchy or centre at all and systems which have a completely centralised organisation. Indeed, the role of secondary agglomerations in site location seems to be slightly stronger than the one of *capita* *civitatum*. 'Super efficient systems' may also be excluded, as they involve a decentralised system of homogeneous spatial distribution, which applies very well to the contemporary Somme, with its extremely homogeneous distribution of rural occupation. 'Primate systems' and their independent peripheries cannot be considered either, because all areas of the Somme are influenced by major centres. Only the term of 'dendritic system' may therefore apply to the Roman Somme: it possesses a major centre which controls minor centres, although in this case there are two major centres, corresponding to the two ancient *civitates*. Despite this seemingly clear-cut classification of the global spatial patterns, strong doubts still persist regarding Region 4 in Eastern Somme, where site location and distribution remains very homogeneous. This picture may change with new archaeological data, but also with new methods of modelling investigation.

4.1.4 The disaggregation of the settlement pattern

In Northern Gaul the Late Roman period began around 260 AD, marking the beginning of the reign of Postumus, who was the founder of the so-called 'Gallic Empire' – a division of the Empire led by military usurpers. It more or less controlled all the provinces of Gaul, Hispania, Germania and Britannia until 274, when the Emperor Aurelian defeated Tetricus, the last 'Gallic Emperor'. From the second half of the 3rd century to the end of the Roman period, Northern Gaul was also subject to other defections and usurpations, as well as raiding by diverse Germanic tribes. In the Somme this was mainly by Saxons and later – from the end of the 4th century onwards – by the Salian Franks, who were the sole authority left in the area by the 5th century.

In archaeological terms, the Late Roman period is characterised by a very important demographic crisis, systematically materialised by the abandonment of sites, especially in the countryside. Paul van Ossel stated that when looking at settlements in a quantitative way, the 4th century settlements display more site location in 'particularly favourable geographical sectors, such as valleys' and 'around certain cities such as Tongres and Cambrai' (Van Ossel 2010, 4). Conversely he noted the massive abandonments in areas where an Early Roman *caput civitatis* was declassed, notably Bavay, which is also evidenced in the vicinity of Cassel and Eu (Van Ossel 2010, 4; Revert

(To be published)). In the case of Bavay, the less fertile soils – soon replaced by the forest of Mormal – compared to the neighbourhood of the city of Cambrai may play an important part in this outcome, whereas around Cassel, the higher maritime influence and the failure of the salt-making industry provide important factors for the near-complete abandonment of the area (Delmaire 2004, 45–47).

This century is also characterised by strong micro-regional variations, with different territories preserving parts of their settlement patterns in specific areas, for example the fertile plains of the Parisian basin persisting much better than the maritime plains (Van Ossel 2010, 4). The architecture and the morphology of rural settlements changes towards smaller scales and very often post-built buildings, although site location is nearly always in continuation of the Early Roman one (Van Ossel 2010, 7–9; Revert 2016, 104–105).

The results do not contradict the traditional view on Late Roman settlement patterns given by Agache, who claimed that the settlements of this period survived better around the main centres, be they towns, secondary agglomerations or sanctuaries (Van Ossel *et al.* 2000, 133; Agache 1978). Wightman also argued that this massive abandonment in rural settlements could be balanced by the apparition of fewer but larger 'villages', correlated with the appearance of large inhumation cemeteries (Wightman 1985, 257–258). Archaeological research, if it has found a few large cemeteries of the sort, for example in Vron in the Somme, has yet to prove that large nucleated villages are linked to these (Joël 1989; Van Ossel *et al.* 2000, 139; Van Ossel 1992). In the Marne valley, the landscape is very similar to that of the Somme and also displays a huge abandonment rate in all areas, even the most fertile ones, already in the 3rd century AD (Van Ossel *et al.* 2000, 139).

Now the methodological issues of the present approach shall be discussed in comparison with other works.

4.2 Methodological issues

Following Kamermans' description of the six main issues in archaeological location models, the methodology employed during this thesis will be discussed in terms of the possible answers it can give through this work, other studies or future explorations. These problems concern the 'quality and quantity of archaeological input data, the relevance of the environmental input data, the lack of temporal and/or spatial resolution, the use of spatial statistics, the testing of predictive models and the need to incorporate social and cultural input data' (Kamermans 2004, 273; Kamermans *et al.* 2004, 6–7). It should be noted that criticism solely given to CRM-oriented models will not be discussed, as this modelling approach tried to follow other inspirations than 'correlative models' which are often discarded as inaccurate and theoretically dubious (Wheatley 2004) and are sometimes used without a scientific critical mindset, which is not required according to local policies in force (van Leusen 2009, 54).

4.2.1 The chronological classification

Contrary to most CRM-oriented practice (Kamermans 2008, 74), explanatory 'deductive ' models must make use of chronological subsets of site data, so as to allow an accurate site location analysis. The approach taken during this thesis does not provide pertinent answers to correct chronological uncertainties. Indeed, by giving the maximal interval described in reports and ceramic inventories, it does not give a very accurate and precise picture of the occupation span of a given settlement phase.

A complete overhaul of the sites' dating is therefore required, as well as an important review of the surveyed sites' ceramics and a qualitative scrutiny of all excavated sites. Though, if such inquiry could be performed, different possible methods could help narrow down chronological uncertainties. Aoristic analysis, for example, consists in attributing weights to specific types of materials and structures in function of their degree of chronological precision: the more precise the chronological classification, the more weight pertains to a given element (Johnson 2004). The assemblages are then summed and the resulting distribution of dated elements acknowledges probabilistic weights, applied on periods standardised by 'dividing the median date of the period (...) by the length of the time span' (Johnson 2004, 451).

Other methods include the interrelation of typological and morphological data with chronological data (Durand-Dastès *et al.* 1998, 161–162): specific morphological settings of settlements may be representative of specific periods. It is also possible to associate surveyed settlement typologies with more precisely dated excavated ones (Nüsslein 2016a; Van der Leeuw *et al.* 2003, 225–238), or by specifying occupation duration per phase in each settlement, which requires very precise excavation data (Favory 2013). This in turn can improve both the chronological precision and accuracy of the classification, as well as better understand and represent the typological evolutions of rural settlements.

If such reclassification of site chronology in the Somme is carried out in the future, it should also expand towards the Late Iron Age, as a strong locational and morphological continuity links it to the Roman period. Only when this is expanded could land-use heritage such as that employed by Nuninger be tested in the Somme (Nuninger *et al.* 2016). A much more detailed and accurate comparison of settlement pattern dynamics through time will also be possible.

4.2.2 The meaning of 'site' and spatial behaviour

When studying spatial and temporal patterns in the countryside – particularly when the dataset is as dependent on survey data as the Somme – one is confronted with difficulties in consistently defining 'sites' and the relations of humans with their landscape. Indeed, a site implicitly represents a unique, entire and indivisible unit, which in this case is then called a rural settlement. Though, neither excavations nor surveys generally give such a complete picture of settlements. On the basis of Kohler and Parker's early studies (Kohler *et al.* 1986, 403), Kamermans noted that the effectiveness of an 'inductive' predictive model is strongly improved by the use of a probabilistic sampling method, i.e. a systematic and controlled survey sample of sites, giving a maximal representation of site data recoverable from the sampled area (Kamermans 2008, 74). Though, just as in the example of the IKAW model of the Netherlands, the Somme was not surveyed in such a systematic manner, as it is impossible to apply on such a large scale. The statistical precision of both studies is therefore very poor, especially if oriented towards a CRM aim, which thankfully is not the case in this thesis.

The meaning of 'site' is very important in the outcome of any predictive model, as indicated by Whitley (2004a, 4). Indeed, the latter noted that evidence for human activity can be lacking – which in this case strongly relates to the lower representation of post-built settlements – and some variables may have a causal relationship with other variables than the sites in themselves, just as some parameters may have been taken unconsciously into account for site location. As an example, the aspect of a cell is clearly a causal variable for cumulative solar radiation, but is not in a causal relation with the rural settlements of the Somme. Furthermore, sites should be divided in functional subsets, as a site can be a burial site, a rural settlement, an urban one, or part of any other subdivision (Kamermans 2008, 74). The subdivision of rural settlements in three meaningful categories, if allowing for the production of slight differences in locational preference, is countered by the fact that other site classifications could very well have produced different patterns.

This approach is still very much akin to an archaeology of place rather than space, as the dependent variable is constituted by artificial settlement centres, indistinguishable in terms of area and shape because they only serve as a validating presence on a specific cell. Furthermore, focal statistics were only allowed in the creation of variables, not during the testing, as it smoothed down all results and the models therefore lost both accuracy and precision. As indicated by van Leusen, even 'cognitive' variables are in fact similar to any other variables used in predictive modelling, for they 'still involve measurable properties of the environment' (Gaffney *et al.* 1995, 370–371)

In reconstructing human behaviours in site location, Whitley determined three types of 'referential variables' to reach these behaviours (Whitley 2004c, 7–8). The direct causal references are variables which were directly perceived by the ancient inhabitants, i.e. derived from their senses. Visibility analysis is part of such reference, but it was not implemented in this thesis due to rural settlements being rarely built with a preference for the few areas that provide vistas in the landscape of the Somme. Indirect causal references are measurable, but not perceived immediately by people's senses. All path distances are part of such proxies to reconstruct human behaviours.

Non-causal references cannot be measured but can be modelled through the use of proxies: for example spatial knowledge could be modelled by combining viewshed analysis and cost-distance to the element whose knowledge is the target of an inquiry (Whitley 2000; Whitley 2002). Non-causal references could have been created in this thesis, using proxies that have already been modelled, but directly referring to these independent variables was preferred. The influence of the unquantifiable social connectivity between the countryside and the urban centres on rural settlement location could nonetheless be inquired, simply by using the proxy of path distances to *capita civitatum*. This is a very crude approximation of such social relations, but one that could indeed have a causal relationship with social connectivity. Land-use heritage also constitutes a non-causal reference, as it is not measurable, but can still be modelled by means of a set of proxies quantifying the temporal length of settlement activity and their closeness (Nuninger *et al.* 2016). All three types of spatial references should be expanded in the future of predictive modelling, but indirect and non-causal references may hold the most influential trove of variables, as they are the ones which relate the most to socioeconomic and cultural decisions in site location.

4.2.3 Evaluation of the models' performance

The method which was designed and applied in this thesis is not a perfect one by any means and now shall be evaluated for its efficiency, advantages and disadvantages, keeping in mind that it was specifically designed for this study case.

First of all, this methodology was aimed at analysing site location preference and is not CRM-oriented. This is a very important starting point because the model would otherwise be deemed insufficient towards cultural resource management. Indeed, basing his judgment on Kvamme's gain (Kvamme 1988, 329), Whitley indicated that the precision and accuracy of any model's high prediction area is sufficient only when its Kvamme's gain is higher than 0.5 (Whitley 2004a, 4). In that sense, only once has such a value been reached in all the models: in the multivariate model of the physical landscape and the distance to archaeological features, in region 3 in the first half of the fourth century (table 21). It should be noted that this poor performance overall could be due to the higher number of classes per model, which was defined differently than the CRM-oriented binary classification (presence or absence). It is still unsure if these models could reach more 'satisfying' statistical efficiencies by changing the way they were built, for in any case, the Somme is an area which does not display extremely radical disparities, both in terms of physical landscape or socio-economic variables. This is very different in more contrasted areas with lower site frequencies, such as the United States.

Whitley himself admitted that models built for site location analysis in homogeneous environments pose many problems: slope and distance to water are not as influential as in near-arid areas, the spatial data may be too inaccurate or auto-correlated, an accurate palaeoenvironment cannot be reconstructed, site datasets are incomplete, the studied areas are of an ‘inappropriate scaling’ and there can be ‘map-edge effects’ (Whitley 2001, 8–9). Regarding these issues, this methodology could build socio-economic variables which were influential for site location, but palaeoenvironmental parameters are still lacking in completeness, accuracy and precision, just as some types of sites are less represented than others. The ‘map-edge effect’ does not have any important influence in the model’s results, as variables were first created on a larger scale than the individual micro-regions.

However, other explanatory, cognitive and theory-based models are sometimes confronted with difficulties in determining site location decisions. Even in areas of extreme environmental heterogeneity such as southern Egypt, Burns *et al.*’s attempts at determining the site location preferences of ancient tombs did not achieve a grading of the only two relevant variables in the model (orientation and ground stability), even working with models of much larger relative gains than those presented in this thesis (up to 0.38) (Burns *et al.* 2008, 5–6).

4.2.4 A non-deterministic landscape?

The results of this thesis clearly show that the Somme is devoid of deterministic parameters in terms of site location, as the multivariate models indicate that rural settlements may be located in any area, even if in different frequencies. Though, with a different protocol and better computational efficiency, a comparison between many potential weighting schemes might have been carried out and some trends may have appeared more clearly elsewhere, rather than only the ones which were presented. Contrary to this assumption, Whitley argues that local variations may defeat any weighting scheme, as individuals would all have a different decision – conscious or not – regarding their environment (Whitley 2004a, 4). Only the global and commonly occurring trends could therefore be showcased by predictive models, as stated by van Leusen (Gaffney *et al.* 1995, 371).

It may be argued that only deterministic parameters may help produce models with high enough statistical gains. Only the reconstruction of the palaeoenvironment may provide such variables, especially through the distribution of ancient soils, but it may also be that any of the Somme's soils could be settled, which remains to be proved or disproved. Thanks to the availability of such data, the IKAW model had a strong focus on the physical landscape, through contemporary soil and geological data (Kamermans 2008, 75). Oppositely, Whitley only recognised socio-economic and cultural variables as having an important cognitive role in site location, whereas he considered most physical environmental parameters as auto-correlated (Whitley 2004a, 8).

The call for cognitive methods professed by Whitley and van Leusen in order to bridge the gap between environmental determinism and theoretical phenomenology is therefore partially concretised by the most recent works in predictive modelling, which put more and more focus on building socio-economic variables. Though, the way GIS works has not changed at all, which implies that the physical and human environments are still measurable classes analysed through place rather than symbolic spaces, which cannot be quantified and evaluated. For example, some 'cognitive' variables may in fact be proxies for an ecological determinism: Binford argues that the distribution of 'critical resources' determines hunter-gatherers' strategies which in turn decide settlement patterns (Binford 1980, 10).

The consideration that 'landscape is an entity that exists by virtue of its being perceived, experienced, and contextualized by people' (Ashmore *et al.* 1999, 1) may be too radical and in fact more of a philosophical and rhetorical concern than an archaeological one. It cannot be denied that there is a physical and non-human element to landscapes, but socio-economic and cultural perceptions of a non-physical or indirectly physical landscape did also play important parts in settlement patterns. The true advance lies in the recognition that both categories of variables are proxies to reconstruct hypotheses on the 'full complexity of archaeological data' (Gaffney *et al.* 1995, 378-379), which can ultimately be verified only through archaeological survey or fieldwork. The fundamental character of complexity permeates the archaeological data in its spatial patterns, temporal aspects and contextual notions, distorted by ‘changes in the natural environment, patterns of activities in the landscape since deposition and research intensity’ (Randsborg 1981, 40).

4.2.5 Accuracy and precision

Finally, the four different ways to evaluate a model’s accuracy and precision (Chi-square test, Kvamme’s gain, indicative value and relative gain), although simple to implement and easy to compare and understand, should not be taken as an absolute and completely reliable assertion of the models, especially because they are theory-based. Indeed, there are many other ways to define a model’s predictive power, but as Whitley noted, an explanatory focus in the modelling process should rely more on accuracy than precision (Whitley 2004b, 315). This is clearly the case in this thesis, where areas where we expect high frequencies of sites are indeed highly predictive, and inversely for areas that expect few sites, but the range of variation between expected and observed sites is generally low.

Kamermans and Wansleeben also noted that the most frequent statistical test method – Chi-square tests –barely ever complies with the two required preconditions: more than 20% of all cells need to constitute a category with an expected frequency of less than 5 and all cells have to have an expected frequency of at least 1 (Kamermans *et al.* 1999, 226; Verhart *et al.* 1992, 100). Only through the careful examination of different evaluating methods and a precise understanding of the model’s variables can such assertion of a model’s effectiveness be reached.

In the future, it would be interesting to test other methods of statistical evaluation on the Roman Somme, such as the probabilistic Attwell and Fletcher test, which can produce probabilities of a category having an important relationship with site distribution (Attwell *et al.* 1987). Bayesian statistics could also be used to define how many sites can be expected in the models’ classes, according to expert judgement and/or data correlation, and with a confidence ratio that could allow further demographic estimates (Finke *et al.* 2008; Verhagen *et al.* 2009, 20; Millard 2005). Dempster-Shafer’s Theory of evidence (DST) may also allow to define uncertainties of the models by attributing a 'belief' value to specific independent variables and their classes, also based on data correlation or expert judgement and generally extending Bayesian statistics in a theoretical ground (Verhagen *et al.* 2011, 583–586; Philip Verhagen *et al.* 2010, 433–434; Ejstrud 2005). But as accuracy is the main concern of the modelling objectives of this thesis, the K_j should be the main target of new statistical testing of the models. This measure is derivative from Kvamme's gain and includes only one additional ratio: the ratio of area without site (Verhart *et al.* 1992; Verhagen 2009, 76). This addition allows a significant filtering and noise-reduction of statistical results.

Ultimately, it can be surmised that the accuracy and precision of a model still depends more on the quantity and quality of its dataset than the statistical methods used. If much effort has been devoted towards improving statistical methods, especially for CRM projects, the analysis of biases in fieldwork data is generally not the main concern of reviewed research. As noted by Verhagen, one should take heed of 'systematic biases that are either the consequence of data collection strategies, or of geological conditions' (Verhagen *et al.* 2009, 22). Despite a few recent attempts at tackling this issue, much effort is still necessary in order to quantify these biases, especially through confidence and representation mapping (Oštir *et al.* 2008; Lietar 2015).

4.2.6 A comparison of different archaeological location models

When comparing this method with other examples of locational models, strong differences are observed in the building methods as well as the results. The site location analysis of the Roman sites of the Island of Brač (Croatia) employed Chi-square tests to identify which univariate models – 8 physical landscape variables – were correlated with site location, as well as the means of each variables' value where a site is located and in the entire landscape, which is useful in order to identify the variations due to continuous variables (especially cost distance and elevation) (Stancic *et al.* 2000, 6–7; Podobnikar *et al.* 2001, 538–540). Surprisingly enough, soils did not seem to have much influence in site location, but this was due to the very low resolution of this variable, just as it would have been in this case. Using multiple linear regression, a multivariate model was created, confirming the influence of a local type of limestone/marble on site location, but otherwise failing to produce statistically significant results. Veljanovski and

Podobnika concluded that an insufficiently precise classification of the Roman sites was the reason for this failure, as some settlements could have had an agricultural function but others not (Stancic *et al.* 2000, 8–9). Then, through cluster analysis of site location preferences, some very small clusters of sites were defined and functional interpretations were given to them (Stancic *et al.* 2000, 9–10). This method could be useful in order to determine the function of settlements for which not enough data is possessed to give one, but the very small number of sites included – 29 'settlements' – in a fairly large area (395 km^2) do not give a truly compelling study case.

This thesis's methodology is in fact much closer to another one also developed on the Island of Brač (Podobnikar *et al.* 2001, 538; Stancic *et al.* 1999), with manually defined variables combined with a Boolean overlay (variables with binary classification) and map algebra, providing a very transparent model, well suited for the very small sample of hillforts included in this case study (8). Social variables were also more prominent in this model, employing distance between hillforts, intervisibility, distance to the sea – thought of as a danger due to piracy and raiders – and distance to barrows (Stancic *et al.* 1999, 234). No specific predictive value was given for this model, but as it only concerns 8 sites, the success of the model could be easily acknowledged. In fact, it could be argued that no testing at all is needed in that case, as the model could be tailored so as to fit the distribution of the hillforts and therefore all explanations regarding their site location were produced during the model building process.

This latter method in fact stems from earlier works in the Croatian Island of Hvar (Gaffney *et al.* 1991, 58–76), where site location analysis was carried out using Chi-square tests on univariate models. When concerned with Protohistorical cairns and tumuli, it helped delimitate possible functions surrounding these structures, while the Roman settlements displayed a very high correlation between positive site location and the best soils and local climates (Gaffney *et al.* 1991, 66–76). Then, the best locations thus determined through univariate Chi-square testing were combined with map algebra and produced an indicative predictive model for Roman settlements. With its non-Boolean classification and overlays, this approach was a strong inspiration for this thesis's models, but the employed variables still have to be created in the Roman Somme.

The IKAW model is the one which is most comparable to this thesis' models, which notably stems from the fact that this example inspired the present methodology towards the application of weighted map algebra in order to build the univariate and multivariate models. As such, the same critics which could be applied to the IKAW also concern this thesis's methodology: any change in the very subjective weights attributed to classes and variables has a very large impact on the model's behaviour (Brandt *et al.* 1992, 271). Though, in its three installments, the IKAW's emphasis on physical landscape variables (contemporary or reconstructed) and binary classification (presence or absence), as well as the CRM-oriented objectives, distinguish this thesis from its methodological inspiration (Deeben 1997; Deeben *et al.* 2002; Deeben 2008).

A final method which could be implemented in order to analyse site location is also a very difficult one to carry out: land-use evaluation/estimate (Kamermans 2000, 133–143). Inspired by ecological and agronomical studies, it first aims to reconstruct ancient landscapes (especially soils), then create conceptual socio-economic models of land-use (agro-pastoral societies, hunter-gatherers, etc); reclassify the ancient landscapes according to their suitability for a specific type of land-use and finally confront the archaeological data and the expected areas where a specific land-use form is expected to be found (Kamermans 2000, 134–135). In this matter, a large quantity of geomorphological and plalynological data is of course necessary, as well as proper research into what requirements are associated with different agro-pastoral traditions. In the case of the Agro Pontino area (Italy), it could help with associating dominant Neolithic agro-pastoral traditions to specific land units, such as permanent farming in densely occupied areas on land units of 'Terracina beach ridges, coastal lagoons and inland lagoons' (Kamermans 2000, 141–142). Its application in South Limburg is less interesting in terms of site location analysis, as no specific variable could be highlighted, but it provided a better CRM-oriented tool than the precedent correlative methods (van Leusen *et al.* 2005, 61–62). This approach was impossible to apply in this thesis's study case, as no such palaeoenvironmental data was available, but future research in the area and neighbouring regions should try to put more emphasis on this land evaluation approach.

Conclusion

This study of the settlement pattern of the Roman Somme, focusing on four similar and yet diverse micro-regions, sought to design, implement and evaluate a relatively simple predictive modelling methodology in order to find out about site location preferences of rural settlements. This study area had never been included in any predictive modelling project and very few studies incorporated hypotheses regarding the definition of specific variables as being influential in site location. Thus, the main research question was: According to site location analysis using predictive modelling, which environmental or human factors influenced the rural settlement patterns of the Somme during the Roman Imperial period? Most of the time, the Somme's physical and human environment was seen as non-influential, in the sense that the surveyed settlement density was so high and widespread that visual inspection of distribution maps did not give relevant hints towards preferential areas or areas forbidding settlement.

Nevertheless, this thesis sought to find out if such assumptions were true or not by confronting a rich but heterogeneous site dataset with variables pertaining to the physical and socioeconomic landscape. The methodology that was designed can be associated with 'deductive' or theory-driven and cognitive predictive models, especially those developed in the Netherlands during the three instalments of the IKAW predictive modelling project. This thesis's models employ multivariate weighted map algebra in a very simple fashion, which does not give the most precise predictive values, but is relatively easy to implement and particularly transparent. Though, as each variable and category's weighting is manually defined through expert judgement, leading this approach to fruition is a lengthy process. Still, another positive aspect of this methodology is its easy acceptance of very diverse independent variables.

Notwithstanding these methodological considerations, the correlation of multiple variables with site location could be evaluated, micro-region by micro-region and throughout time. While the physical landscape does play an important role in site location – especially slopes, landforms and the path distance to rivers – other variables such as aspect and solar radiation were deemed irrelevant variables. The socio-economic variables which were modelled were concerned with the relative path distance to archaeological features of the economic landscape: cities, secondary agglomerations, sanctuaries and Roman roads. All of these displayed a higher correlation with site location, especially secondary agglomerations and sanctuaries.

The rural settlement pattern and its evolution globally display a non-deterministic relation with its landscape: all areas could be settled, albeit in different frequencies. Be it in the beginning of the 2nd century AD or the Late Roman period, rural settlements were generally located in relatively flat or gently sloping areas and were very well situated in regards to socio-economic focal points. If post-built settlements were slightly more often located in areas on the margins of socio-economic foci and villas were closer to them, they were both following the general trend of the main category of rural sites: stone-built settlements. The first three micro-regions – all contained in the *civitas ambianorum* – behaved similarly in regards to the predictive models, but the fourth micro-region, in the western part of the *civitas viromanduorum*, generally showed a very homogeneous site distribution, which could be caused by the very flat and favourable physical landscape, as well as the subsequently eased accessibility of socio-economic features.

Both the diversity of independent variables and the robustness of the methodology need radical extensions and statistical improvements so as to be more reliable and fully replicable in other areas. New hypotheses on human behaviours playing a role in site location and therefore new socioeconomic and cultural variables should also be tested or developed, acknowledging a more complete set of human decisions regarding their rural settlements. Reconstructions of palaeoenvironmental variables should also be the focus of much effort, as they provide more accurate depictions of past-landscapes. The incorporation of probabilistic statistics in the model will also improve the representation and accuracy of the site data and its chronological classification, on which any model's results are immensely dependent. Many approaches exist to build and test models and a comparison of those on the same datasets could prove enlightening regarding the suitability of a given methodology in the perspective of site location analysis. Expanding the geographical and chronological boundaries of this study would help put this thesis's results and new methods into a wider perspective.

Finally, archaeological location predictive models which incorporate socio-economic and cultural variables as well as accurate variables of the physical environment are very useful tools to dissect settlement patterns throughout time and space. They also constitute a bridge between traditional landscape archaeology and behavioural landscape archaeology, as they can quantify and evaluate scientific hypotheses on archaeological site location preferences.

Bibliography

Agache, R., 1975. La campagne à l'époque romaine dans les grandes plaines du nord de la France a d'après les photographies aériennes. *Aufstieg und Niedergang der Römischen Welt* 2(4), 658–713.

Agache, R., 1976. Les fermes indigènes d'époque pré-romaine et romaine dans le bassin de la Somme. *Cahiers archéologiques de Picardie* 3(1), 117–138.

Agache, R., 1978. *La Somme pré-romaine et romaine: d'après les prospections aériennes à basse altitude*. Amiens: Société des Antiquaires de Picardie.

Agache, R., 1997. Repérage des sanctuaires gaulois et gallo-romains dans les campagnes du bassin de la Somme et ses abords (information). *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 141(2), 551–566.

Agache, R., F. Vasselle and E. Will, 1965. Les villas gallo-romaines de la Somme : aperçu préliminaire. *Revue du Nord* 47(187), 541–576.

Brunaux, J.L. (ed), 1999. Ribemont-sur-Ancre (Somme): Bilan préliminaire et nouvelles hypothèses. *Gallia* 56(1), 177–283.

Ashmore, W. and A.B. Knapp, 1999. *Archaeologies of landscape: contemporary perspectives*. Malden, Mass: Blackwell.

Attwell, M.R. and M. Fletcher, 1987. An analytical technique for investigating spatial relationships. *Journal of Archaeological Science* 14(1), 1–11.

Bayard, D., 1993. Sépultures et villae en Picardie au Haut-Empire : quelques données récentes. *Supplément à la Revue archéologique du centre de la France* 6(1), 69–80.

Bayard, D., N. Buchez and P. Depaepe (eds), 2011. *Quinze ans d'archéologie préventive sur les grands tracés linéaires en Picardie. Première partie*. Amiens: Université de Picardie.

Bayard, D., N. Buchez and P. Depaepe (eds), 2014. *Quinze ans d'archéologie préventive sur les grands tracés linéaires en Picardie: Seconde partie*. Amiens: Université de Picardie.

Bayard, D. and J.L. Collart (eds), 1996. *De la ferme indigène à la villa romaine: la romanisation des campagnes de la Gaule ; actes du deuxième Colloque de l'Association AGER tenu à Amiens (somme) du 23 au 25 septembre 1993*. Amiens: Service régional de l'archéologie en Picardie (Revue archéologique de Picardie, Numéro spécial 11).

Bayard, D. and J.L. Collart, 1996. De la ferme indigène à la villa romaine. La romanisation des campagnes de la Gaule, in D. Bayard and J.L. Collart (eds), *De la ferme indigène à la villa romaine: la romanisation des campagnes de la Gaule ; actes du deuxième Colloque de l'Association AGER tenu à Amiens (somme) du 23 au 25 septembre 1993*. Amiens: Service régional de l'archéologie en Picardie (Revue archéologique de Picardie, Numéro spécial 11), 5–8.

Ben Redjeb, T., L. Duvette, P. Quérel, C. Chaidron, S. Lepetz, V. Matterne and D. Maréchal, 2005. Les campagnes antiques : bilans et perspectives. *Revue archéologique de Picardie* 3(1), 177–222.

Ben Redjeb, T., 2013. *Carte archéologique de la Gaule: 80/2. La Somme*. Paris: Éditions de la Maison des sciences de l'homme.

Bender, B., 1999. *Stonehenge: Making Space*. Oxford, New York: Berg 3PL.

Bertoncello, F., É., Fovet, C. Gandini, F. Trément and L. Nuninger, 2008. The spatio-temporal dynamics of settlement patterns from 800 BC to 800 AD, in Central and Southern Gaul, in C. Gandini, F. Favory and L. Nuninger (eds), *7 millennia of territorial dynamics : settlement pattern, production and trades from Neolithic to Middle Ages: closing colloquium of The ArchaeDyn program ('Spatial dynamics of settlement patterns and natural resources: towards an integrated analysis over a long term, from Prehistory to the Middle Ages'), June 2008, Dijon, France*. Oxford: Archaeopress, 141–148.

Binford, L.R., 1980. Willow Smoke and Dogs? Tails: Hunter-Gatherer Settlement Systems and Archaeological Site Formation. *American Antiquity* 45(1), 4–20.

Blondiau, L., 2014. L'occupation rurale gallo-romaine de la vallée de la Selle: première approche, in M. Reddé (ed), *Séminaire Workshop 1 : Établissements ruraux du Nord et de l'Est de la Gaule: données récentes*. Paris: Ecole Pratique des Hautes Etudes, 1–15.

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.

Briquet, A., 1930. *Le littoral du nord de la France et son évolution morphologique ; suivi d'un appendice: L'évolution du rivage du nord de la France et l'activité de l'homme*. Paris: Armand Colin.

Burgess, R.J., K.L. Kvamme, P.R. Nickens, A.D. Reed and G.C. Tucker J., 1980. *Class II Cultural Resource Inventory of the Glenwood Springs Resource Area Grand Junction District, Colorado Part II Predictive Model of Archaeological Site Location in the Glenwood Springs Resource Area*. Montrose: Nickens and Associates.

Burns, G., A.K. Fronabarger and T.G. Whitley, 2008. Predictive modeling of cultural resources in the Theban Necropolis, Luxor, Egypt, in A.G. Posluschny, K. Lambers and I. Herzog (eds), *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, 2.-6. April 2007*. Bonn: Dr. Rudolf Habelt GmbH (Kolloquien zur Vor- und Frühgeschichte 10), 1–7.

Collart, J.-L., 1996. La naissance de la villa en Picardie : la ferme gallo-romaine précoce, in D. Bayard and J.L. Collart (eds), *De la ferme indigène à la villa romaine: la romanisation des campagnes de la Gaule ; actes du deuxième Colloque de l'Association AGER tenu à Amiens (somme) du 23 au 25 septembre 1993*. Amiens: Service régional de l'archéologie en Picardie (Revue archéologique de Picardie, Numéro spécial 11), 121–156.

Collart, J.-L., J.O. Guilhot, R. Hanoune and M. Dondin-Payre, 2004. *La marque de Rome : Samarobriva et les villes du nord de la Gaule*. Amiens: Musée de Picardie.

Curry, M., 1998. *Digital Places: Living with Geographic Information Technologies*. London: Routledge.

De Reu, J., J. Bourgeois, P. De Smedt, A. Zwertvaegher, M. Atrop, M. Bats, P. De Maeyer, P. Finke, M. Van Meirvenne, J. Verniers and P. Crombé, 2011. Measuring the relative topographic position of archaeological sites in the landscape, a case study on the Bronze Age barrows in northwest Belgium. *Journal of Archaeological Science* 38(12), 3435–3446.

De Reu, J., J. Bourgeois, P. De Smedt, A. Zwertvaegher, V. Gelorini, C. Wei, M. Atrop, M. Bats, P. De Maeyer, P. Finke, M. Van Meirvenne, J. Verniers and P. Crombé, 2013. Application of the topographic position index to heterogeneous landscapes. *Geomorphology* 186, 39–49.

Deeben, J., D.P. Hallewas and T.J. Maarleveld, 2002. *Predictive modelling in archaeological heritage management of the Netherlands: the indicative map of archaeological values (2nd generation)*. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek.

Deeben, J.H.C., 1997. Beyond the crystal ball: predictive modelling as a tool in archaeological heritage management and occupation history, in W. J. H. Willems, H. Kars and D.P. Hallewas (eds), *Archaeological Heritage Management in the Netherlands: Fifty Years State Service for Archaeological Investigations*. Assen, Amersfoort: Van Gorcum, Rijksdienst voor het Oudheidkundig Bodemonderzoek, 76–118.

Deeben, J.H.C. (ed), 2008. *De Indicatieve Kaart van Archeologische Waarden, derde generatie*. Amersfoort: Rijksdienst voor Archeologie, Cultuurlandschap en Monumenten.

Delmaire, R., 2004. Permanences et changements des chefs-lieux de cités au Bas-Empire: l'exemple du nord-ouest de la Gaule Belgique, in A. Ferdière (ed), *Capitales éphémères. Des capitales de cités perdent leur statut dans l'Antiquité tardive, Actes du colloque Tours 6-8 mars 2003*. Tours: Fédération pour l'édition de la Revue archéologique du Centre de la France (Supplément à la Revue archéologique du centre de la France 25), 39–50.

Deru, X., 2012. L'habitat au sein des agglomérations du nord de la Gaule, in L. Akovleva, A. Korvin-Piotrovski and F. Djindjian (eds), *L'archéologie du bâti en Europe, Actes du congrès archéologique franco-ukrainien, Paris EHESS 1-3 octobre 2009*. Kiev: Korvin Press, Chap. 14.

Durand-Dastès, F., F. Favory, J.L. Fiches, and S. Robert, 1998. *Des oppida aux métropoles: archéologues et géographes en vallée du Rhône*. Paris: Anthropos.

Duvette, L., 2008. *Plateforme aéroindustrielle de Méaulte. Site 3. Document final de synthèse*. Amiens: INRAP.

Ejstrud, B., 2005. Taphonomic Models: Using Dempster-Shafer theory to assess the quality of archaeological data and indicative models, in M. van Leusen and H. Kamermans (eds), *Predictive modelling for archaeological heritage management: a research agenda*. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek (Nederlandse Archeologische Rapporten 29), 183–194.

Ervynck, A., C. Baeteman, H. Demiddele, Y. Hollevoet, M. Pieters, J. Schelvis, D. Tys, M. Van Strydonck and F. Verhaeghe, 1999. Human occupation because of a regression, or the cause of a transgression? A critical review of the interaction between geological events and human occupation in the Belgian coastal plain during the first millenium AD. *Probleme der Küstenforschung im südlichen Nordseegebiet* 26, 97–121.

Fauduet, I., 2010. *Les temples de tradition celtique*. Paris: Errance.

Fauduet, I., D. Bertin and V. Rey, 1993. *Atlas des sanctuaires romano-celtiques de Gaule*. Paris: Errance.

Favory, F., L. Nuninger, F. Bertoncello, E. Fovet, E. Gauthier, N. Poirier, M. George-Leroy, K. Ostir, L. Saligny, F. Trément and O. Weller, 2008. *ArchaeDyn. Dynamique spatiale du* *peuplement et ressources naturelles: vers une analyse intégrée dans le long terme, de la Préhistoire au Moyen Âge. Rapport de fin de contrat ACI Espaces et territoires (contrat ET28, 2005-2007)*. Besançon: Université de Franche-Comté.

Favory, F., 2012. La dure condition des agglomérations secondaires. *Les Nouvelles de l'archéologie* 127, 40–44.

Favory, F., 2013. *Archaedyn: Dynamique spatiale des territoires de la Préhistoire au Moyen Âge, Programme BLANC 2008 (Compte rendu de fin de projet)*. Paris: Agence Nationale de la Recherche.

Favory, F. and L. Nuninger, 2008. Archaedyn 2005-2007 : ambitions, achievements and accomplishments, in *7 millennia of territorial dynamics : settlement pattern, production and trades from Neolithic to Middle Ages: closing colloquium of The ArchaeDyn program ('Spatial dynamics of settlement patterns and natural resources: towards an integrated analysis over a long term, from Prehistory to the Middle Ages'), June 2008, Dijon, France*. Oxford: British Archaeological Reports (BAR international series 2370), 235–243.

Favory, F. and L. Nuninger, 2012. Introduction to the Archaedyn project. Spatial dynamics of settlement and natural ressources : toward un integrated analysis over the long terme from Prehistory to Middle Ages, in C. Gandini, F. Favory and L. Nuninger (eds), *7 millennia of territorial dynamics : settlement pattern, production and trades from Neolithic to Middle Ages: closing colloquium of The ArchaeDyn program ('Spatial dynamics of settlement patterns and natural resources: towards an integrated analysis over a long term, from Prehistory to the Middle Ages'), June 2008, Dijon, France*. Oxford: British Archaeological Reports (BAR international series 2370), 21–22.

Ferdière, A., C. Gandini, P. Nouvel, and J.L. Collart, 2010. Les grandes villae «à pavillons multiples alignés» dans les provinces des Gaules et des Germanies: répartition, origine et fonctions. *Revue archéologique de l'Est* 59(2), 357–446.

Fichtl, S., 2009. La villa gallo-romaine, un modèle gaulois? Réflexions sur un plan canonique, in S. Rieckhoff-Pauli, S. Grunwald, J. K. Koch and D. Mölders (eds), *Festschrift für Sabine Rieckhoff zum 65. Geburtstag*. Bonn: Habelt (Professur für Ur-und Frühgeschichte der Universität Leipzig), 439–448.

Finke, P.A., E. Meylemans and J. Van de Wauw, 2008. Mapping the possible occurrence of archaeological sites by Bayesian inference. *Journal of Archaeological Science* 35(10), 2786–2796.

Fleming, A., 2006. Post-processual Landscape Archaeology: a Critique. *Cambridge Archaeological Journal* 16(3), 267.

Gaffney, V. and M. van Leusen, 1995. Postscript - GIS, environmental determinism and archaeology: a parallel text, in G. R. Lock and Z. Stančič (eds), *Archaeology and Geographical Information Systems: A European Perspective* London, Bristol: Taylor and Francis, 367–382.

Gaffney, V. and Z. Stancic, 1991. *GIS Approaches to Regional Analysis: Case Study of the Island of Hvar*. Ljubljana: University of Ljubljana, Research Institute of the Faculty of Arts and Science.

Gandouin, E., B. Van Vliet-Lanoë, E. Franquet, V. Andrieu-Ponel, D.H. Keen, P. Ponel, M. Meurisse, J. Bruhlet and M. Brocandel, 2007. Analyse en haute résolution de l'enregistrement de la transgression holocène dans un secteur subsident du littoral français : le bassin-marais de Saint-Omer (Pas-de-Calais, France). *Géologie de la France* 1, 11–32.

Garnsey, P., 1996. *Famine et approvisionnement dans le monde gréco-romain. Réactions aux risques et aux crises*. Paris: Les Belles Lettres.

Grewe, K., 2004. Alle Wege führen nach Rom - Römerstraßen im Rheinland und anderswo, in H. Koschik and K. Grewe (eds), *'Alle Wege führen nach Rom ...' Internationales Römerstraßenkolloqium Bonn (vom 25. bis 27. Juni 2003)*. Pulheim: Rhein Eifel Mosel-Verlag (Materialien zur Bodendenkmalpflege im Rheinland 16), 9–42.

Groenhuijzen, M.R. and P. Verhagen, 2015. Exploring the dynamics of transport in the Dutch limes. *eTopoi. Journal for Ancient Studies* 1.

Güimil-Fariña, A. and C. Parcero-Oubiña, 2015. 'Dotting the joins': a non-reconstructive use of Least Cost Paths to approach ancient roads. The case of the Roman roads in the NW Iberian Peninsula. *Journal of Archaeological Science* 54, 31–44.

Habermehl, D., 2011. Exploring villa development in the northern provinces of the Roman empire, in N. Roymans and T. Derks (eds), *Villa landscape in the Roman north. Economy, culture and lifestyles*. Amsterdam: Amsterdam University Press (Amsterdam Archaeological Studies 17), 64–83.

Habermehl, D., 2013. *Settling in a Changing World: Villa Development in the Northern Provinces of the Roman Empire*. Amsterdam: Amsterdam University Press.

Harris, T. and G. Lock, 2006. Enhancing Predictive Archaeological Modeling, in M. W. Mehrer and K. L. Wescott (eds), *GIS and Archaeological Site Location Modeling*. Boca Raton: CRC Press, 41–62.

Herzog, I., 2010. Theory and Practice of Cost Functions, in F. Contreras, M. Farjas and F.J. Melero (eds), *Fusion of Cultures. Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain, April 2010*. Oxford: Archaeopress (BAR international series 2494), 375–382.

Hodder, I., 1992. *Theory and practice in archaeology*. London: Routledge.

Jochim, M.A., 1976[1]. *Hunter-gatherer Subsistence and Settlement: a Predictive Model*. New York: Academic Press Inc.

Joël, B., 1989. Evolution d'une population d'immigrés en Gaule du nord de la fin du IVe siècle à la fin du VIIe siècle : Vron (Somme). *Dossier de documentation archéologique* 13, 57–71.

Johnson, I., 2004. Aoristic analysis: seeds of a new approach to mapping archaeological distributions through time, in K. Fischer-Ausserer, W. Börner, M. Goriany and L. Karlhuber-Vöckl (eds), *[Enter the Past] The E-way into the Four Dimensions of Cultural Heritage. CAA 2003 I Computer Applications and Quantitative Methods in Archaeology I Proceedings of the 31st Conference, Vienna, Austria, April 2003*. Oxford: Archaeopress (BAR international series 1227), 448–452.

Joyce, J. and P. Verhagen, 2016. Simulating the Farm: Computational Modelling of Cattle and Sheep Herd Dynamics for the Analysis of Past Animal Husbandry Practices, in R. Hermans and A. de Rooy (eds), *3rd International Landscape Archaeology Conference 2014 Rome, 17th – 20th of September 2014*, 1–17. Amsterdam: VU E-Publishing.

Judge, W.J. and L. Sebastian (eds)., 1988. *Quantifying the present and predicting the past : theory, method, and application of archaeological predictive modeling*. Denver, Washington: United States Bureau of Land Management.

Kamermans, H., 2000. Land evaluation as predictive modelling: a deductive approach, in G. R. Lock (ed), *Beyond the Map: Archaeology and Spatial Technologies. Proceedings of the NATO Advanced Workshop, Ravello, Italy, 1-2 October 1999*. Amsterdam: IOS Press (Series A: Life Sciences 321), 124–146.

Kamermans, H., J. Deeben, D.P. Hallewas, M. van Leusen, P. Verhagen, P. Zoetbrood, K. Fischer-Ausserer, W. Börner, M. Goriany and L. Karlhuber-Vöckl, 2004. Deconstructing the crystal ball: the state of the art of predictive modelling for archaeological heritage management in the Netherlands, in K. Fischer-Ausserer, W. Börner, M. Goriany and L. Karlhuber-Vöckl (eds), *[Enter the Past] The E-way into the Four Dimensions of Cultural Heritage. CAA 2003 I Computer Applications and Quantitative Methods in Archaeology I Proceedings of the 31st Conference, Vienna, Austria, April 2003*. Oxford: Archaeopress (BAR international series 1227), 1–25.

Kamermans, H., 2004. The Application of Predictive Modelling in Archaeology: Problems and Possibilities, in F. Nicolucci and S. Hermon (eds), *Beyond the artefact. Digital Interpretation of the Past. Proceedings of CAA2004. Prato 13-17 April 2004*. Budapest: Archaeolingua, 273–277.

Kamermans, H., 2008. Smashing the Crystal Ball. A Critical Evaluation of the Dutch National Archaeological Predictive Model (IKAW). *International Journal of Humanities and Arts Computing* 1(1), 71–84.

Kamermans, H. and M. Wansleeben, 1999. Predictive modelling in Dutch archaeology, joining forces, in J. A. Barceló, I. Britz and A. Vila (eds), *New Techniques for Old Times. CAA98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998*. Oxford: Archaeopress (BAR international series 757), 225–229.

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., 1983. Computer processing techniques for regional modeling of archaeological site locations. *Advances in Computer Archaeology* 1(1), 26–52.

Kvamme, K.L., 1988. Development and testing of quantitative models, in L. Sebastian and W. J. Judge (eds), *Quantifying the present and predicting the past: theory, method, and application of archeological predictive modeling*, Denver, Washington: United States Bureau of Land Management, 325-427.

Kvamme, K.L., 2006. There and Back Again, in M. W. Mehrer and K. L. Wescott (eds), *GIS and Archaeological Site Location Modeling*. Boca Raton: CRC Press, 3–38.

Kvamme, K.L., 2012. Spatial Information Technologies and Landscape Archaeology: Past Problems and Future Directions, in W. Bebermeier, R. Hebenstreit and J. Krause (eds), *Landscape Archaeology. Proceedings of the International Conference Held in Berlin, 6th – 8th June 2012*. Berlin: Excellence Cluster Topoi (Special Volume 3), 335–340.

Le Glay, M., 1975. La Gaule romanisée, in G. Duby and A. Wallon (eds), *Histoire de la France rurale. Tome 1: La formation des campagnes françaises des origines au XIVe siècle*. Paris: Editions du Seuil (L'Univers Historique), 191–285.

Lemaire, F., 1994. *Pont-de-Metz (Somme), le Champ aux Oisons I, le Champ aux Oisons II, le Champ Pillard*. Amiens: Service Régional de l'Archéologie.

Lemaire, F. and P. Rossignol, 1996. Un exemple exceptionnel d'établissement agricole romain précoce à Conchil-le-Temple ' le Fond de la Commanderie ' (Pas-de-Calais). Résultats préliminaires, in D. Bayard and J.L. Collart (eds), *De la ferme indigène à la villa romaine: la romanisation des campagnes de la Gaule ; actes du deuxième Colloque de l'Association AGER tenu à Amiens (somme) du 23 au 25 septembre 1993*. Amiens: Service régional de l'archéologie en Picardie (Revue archéologique de Picardie, Numéro spécial 11), 185–202.

Lepetz, S. and Matterne, V., 2003. Élevage et agriculture dans le Nord de la Gaule durant l'époque gallo-romaine : une confrontation des données archéologiques et carpologiques. *Revue archéologique de Picardie* 1(1), 23–35.

van Leusen, M., J.H.C. Deeben, D.P. Hallewas, P. Zoetbrood, H. Kamermans and P. Verhagen, 2005. A Baseline for Predictive Modelling in the Netherlands, in M. van Leusen and H. Kamermans (eds), *Predictive modelling for archaeological heritage management: a research agenda*. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek (Nederlandse Archeologische Rapporten 29), 25–92.

van Leusen, M., 2009. Archaeological predictions contested: the role of the Dutch Indicative Map of Archaeological Values (IKAW) in local planning procedures, in H. Kamermans, M. van Leusen and P. Verhagen (eds), *Archaeological Prediction and Risk Management. Alternatives to Current Practice*. Leiden: Leiden University Press (Archaeological Studies Leiden University 17), 49–62.

Leveau, P. (ed), 2009. *Les formes de l'habitat gallo-romain: terminologies et typologies à l'épreuve des réalités archéologiques*. Pessac: Fédération Aquitania Publications.

Leveau, P., P. Gros, and F. Trément, 1999. La recherche sur les élites gallo-romaines et le problème de la villa, in A. Antoine (ed), *Campagnes de l'Ouest: Stratigraphie et relations sociales dans l'histoire*. Rennes: Presses universitaires de Rennes, 287–302.

Lietar, C., 2015. Cartes de représentativité et de confiance: une application à l'échelle de l'Île-de-France au néolithique moyen, in J. Brancier, C. Rémeaud and T. Vallette (eds), *Des vestiges aux sociétés: Regards croisés sur le passage des données archéologiques à la société sous-jacente. Archéo.Doct 7*. Paris: Publications de la Sorbonne (Archéo.Doct 7), 213–230.

Lock, G. and J. Pouncett, 2010. Walking the Ridgeway Revisited: The Methodological and Theoretical Implications of Scale Dependency for the Derivation of Slope and the Calculation of Least-Cost Pathways, in B. Frischer, J. Webb Crawford and D. Koller (eds), *Making History Interactive. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 37th International Conference, Williamsburg, Virginia, United States of America, March 22-26 2009*. Oxford: Archaeopress (BAR international series 757), 192–203.

Lock, G.R., 2001. Theorising the practice or practicing the theory: archaeology and GIS. *Archaeologia Polona* 39, 153–164.

Mantel, E., S. Dubois, S. and S. Devillers, 2006. Une agglomération antique sort de l'anonymat (Eu, 'Bois L'Abbé', Seine-Maritime) : Briga ressuscité. *Revue archéologique de Picardie* 3(1), 31–50.

Matterne, V., 2001. *Agriculture et alimentation végétale durant l'âge du Fer et l'époque gallo-romaine en France septentrionale*. Autun: Editions Mergoil.

Meurisse, M., B. Van Vliet-Lanoë, B. Talon and P. Recourt, 2005. Complexes dunaires et tourbeux holocènes du littoral du Nord de la France. *Comptes Rendus Geoscience* 337(7), 675–684.

Millard, A., 2005. What can Bayesian statistics do for archaeological predictive modelling ? In M. van Leusen and H. Kamermans (eds), *Predictive modelling for archaeological heritage management: a research agenda*. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek (Nederlandse Archeologische Rapporten 29), 169–181.

Morphy, H., 1993. Colonialism, history and the construction of place: the politics of landscape in northern Australia, in B. Bender (ed), *Landscape: politics and perspectives.* New York: Berg, 205–243.

Morris, E.A., B.J. Lutz, N.T. Ohr, T.J. Kloberdanz, K.L Kvamme and C. Pool, 1975. *Archaeological Survey of the Narrows Unit Project Morgan and Weld Counties, Northeastern Colorado*. Lincoln: Midwest Archaeological Center.

Morris, E.A., K.L. Kvamme, N.T. Ohr, M.D. Metcalf, and H.M. Davidson, 1979. *Archeology of the Boxelder Project: a Water Control Project in Larimer County North Central Colorado, 1972-1979 (and Addendum)*. Denver: Colorado Heritage Center.

Nuninger, L., 2002. *Peuplement et territoires protohistoriques du VIIIe au Ier s. av. J.-C. en Languedoc oriental (Gard-Hérault)*. Besançon (unpublished Ph.D. thesis Université de Franche-Comté).

Nuninger, L., F.P. Tourneux, and F. Favory, 2008. From Archaeomedes to Archaedyn, in A. G. Posluschny, K. Lambers and I. Herzog (eds), *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, 2.-6. April 2007*. Bonn: Dr. Rudolf Habelt GmbH (Kolloquien zur Vor- und Frühgeschichte 10), 278.

Nuninger, L., P. Verhagen, F. Bertoncello, and A. Castrorao Barba, 2016. Estimating 'Land Use Heritage' to Model Changes in Archaeological Settlement Patterns, in R. Hermans and A. de Rooy (eds), *3rd International Landscape Archaeology Conference 2014 Rome, 17th – 20th of September 2014.* Amsterdam: VU E-Publishing, 1–12.

Nuninger, L., P. Verhagen, F.P. Tourneux, F. Bertoncello, and K. Jeneson, 2012. Contextes spatiaux et transformation du système de peuplement : approche comparative et prédictive, in F. Bertoncello and F. Braemer (eds), *XXXIIes Rencontres Internationales d'Archéologie et d'Histoire d'Antibes : « Variabilité environnementales, mutations sociales. Nature, intensité, échelles et temporalité des changements. » Actes des rencontres 20-22 octobre 2011*. Antibes, Nice: Editions de l'APDCA, CEPAM, 139–154.

Nüsslein, A., 2016a. Nuancer et améliorer l'analyse des systèmes de peuplement basée sur les données de prospections pédestres: l'étude de la dynamique d'occupation antique du Plateau lorrain. *Archeologia e Calcolatori* 27, 47–64.

Nüsslein, A., 2016b. *Les campagnes entre Moselle et Rhin dans l'Antiquité. Analyse comparative des dynamiques spatiales et temporelles du peuplement de quatre micro-régions du I er s. avant J-C. au V e s. après J-C.* Strasbourg (unpublished Ph.D. thesis Université de Strasbourg).

Oštir, K., Kokalj, Z. and Nuninger, L., 2008. Confidence maps: a tool to evaluate archaeological data's relevance in spatial analysis, in A. G. Posluschny, K. Lambers and I. Herzog (eds), *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, 2.-6. April 2007*. Bonn: Dr. Rudolf Habelt GmbH (Kolloquien zur Vor- und Frühgeschichte 10), 272–277.

Ouzoulias, P., 2006. *L'économie agraire de la Gaule : aperçus historiographiques et perspectives archéologiques*. Besançon (unpublished Ph.D. thesis Université de Franche-Comté).

Ouzoulias, P., 2009. Place et rôle de la petite exploitation agricole dans la gaule romaine : un débat en cours. *Revue archéologique* 1, 149–155.

Ouzoulias, P., 2011. La villa dans l'Est des Gaules. Un témoin de la romanisation? In M. Reddé, P. Barral, F. Favory, J.P. Guillaumet, M. Joly, J.Y. Marc, P. Nouvel, L. Nuninger and C. Petit (eds), *Aspects de la romanisation dans l'Est de la Gaule*. Glux-en-Glenne: Bibracte Centre Archéologique Européen (Bibracte 21), 475–486.

Ouzoulias, P., 2014. ' Nos natura non sustinet': à propos de l'intensification agricole dans quatre terroirs du nord des Gaules. *Gallia* 71(2), 307–328.

Ouzoulias, P. and P. Van Ossel, 2009. Petites et grandes exploitations agricoles : le cas de la Plaine de France, in P. Leveau, C. Raynaud, R. Sablayrolles and F. Trément (eds), *Les formes de l'habitat rural gallo-romain. Terminologie à l'épreuve des réalités archéologiques, actes du colloque Ager VIII, Toulouse 2007*. Bordeaux: Aquitania (Aquitania supplément 17), 111–121.

Pannetier, V., 1996. Les parties agricoles des villae dans le Nord de la Gaule (Ier au IIIe siècle après J.-C). *Revue du Nord* 78(318), 163–175.

Petit, J.-P. (ed), 2000. *Le complexe des thermes de Bliesbruck (Moselle). Un quartier public au cœur d'une agglomération secondaire de la Gaule Belgique*. Paris, Metz: Exé Productions, Editions Serpenoise.

Petit, J.-P. and M. Mangin, 1994. *Atlas des agglomérations secondaires de la Gaule Belgique et des Germanies*. Paris: Editions Errance.

Pichon, B., 2006. Les villes gallo-romaines de la cité des Ambiens (Amiens exclue). *Bulletin de la Société des Antiquaires de Picardie* 170(677–678), 527–539.

Pichon, B., 2009. Formes et rythmes de la romanisation dans l'ouest de la Gaule Belgique. *Pallas. Revue d'études antiques* 80, 317–350.

Pichon, B., 2012. Le destin des agglomérations secondaires urbaines de l'ouest de la Gaule Belgique aux IIIe et IVe siècles, in M. Cavalieri (ed), *Industria apium : l'archéologie : une démarche singulière, des pratiques multiples : hommages à Raymond Brulet*. Louvain: Presses Universitaires de Louvain (FERVET OPVS 1), 123–134.

Pickering, M., 2003. *Modelling hunter-gatherer settlement patterns: an Australian case study*. Oxford: Archaeopress.

Podobnikar, T., T. Veljanovski, Z. Stancic, Z. and K. Ostir, 2001. Archaeological Predictive Modelling in Cultural Resource Management, in M. Konecny (ed), *GI in Europe: Integrative - Interoperable - Interactive, Proceedings of the 4th Agile Conference on Geographic Information Science in Europe Brno, 19-21 April 2001*. Brno: Masaryk University Brno, Citeseer, 535–544.

Quérel, P. and M. Feugère, 2000. *L'établissement rural antique de Dury (Somme) et son dépôt de bronze, IIIe s. av. J.-C.-IVe apr. J.-C.* Villeneuve-d'Ascq: Revue du Nord (Revue du Nord hors-série, Collection Art et Archéologie).

Raepsaet, G., 2002. *Attelages et techniques de transport dans le monde gréco-romain*. Brussels: Livre Timperman.

Randsborg, K., 1981. Complexity, Archaeological Data, Social Equalities and Cultural Diversity, in S. Van der Leeuw (ed), *Archaeological approaches to the study of social complexity*. Amsterdam: Amsterdam University Press (Cingula 6), 38–53.

Reddé, M. (ed), 2016. *Méthodes d'analyse des différents paysages ruraux dans le nord-est de la Gaule romaine*. Paris: RurLand.

Revert, N. (To be published). Entre continuité et ruptures: L'occupation littorale en Gaule Belgique seconde durant l'Antiquité Tardive, in D. Moreau, E. Deshoux and A. Gauthier (eds), *Deuxième colloque étudiant en Sciences Historiques, Université de Lille SHS*, Villeneuve-d'Ascq: Université de Lille.

Revert, N., 2016. *L'occupation du littoral de la Manche continentale au Bas-Empire*. Villeneuve-d'Ascq (unpublished MSc thesis Université de Lille SHS).

Romanowska, I., 2015. So You Think You Can Model? A Guide to Building and Evaluating Archaeological Simulation Models of Dispersals. *Human Biology* 87(3), 169.

Roth, J., 1999. *The logistics of the Roman army at war (264 B.C.-A.D. 235)*. Leiden, Boston: Brill.

Roymans, N. and T. Derks (eds), 2011. *Villa landscapes in the Roman north: economy, culture, and life-styles*. Amsterdam: Amsterdam University Press.

Roymans, N., T. Derks and H. Hiddink, 2015. *The Roman Villa of Hoogeloon and the Archaeology of the Periphery*. Amsterdam: Amsterdam University Press.

Sommé, J., 1977. *Les Plaines du Nord de la France et leur bordure : étude géomorphologique*. Paris: Diffusion Honoré Champion.

Stancic, Z. and K.L. Kvamme, 1999. Settlement pattern modelling through Boolean overlays of social and environmental variables, in J. A. Barceló, I. Briz and A. Vila (eds) *New Techniques for Old Times. CAA98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998*. Oxford: Archaeopress (BAR international series 757), 231–238.

Stancic, Z. and T. Veljanovski, 2000. Understanding Roman settlement patterns through multivariate statistics and predictive modelling, in G. R. Lock (ed), *Beyond the Map: Archaeology and Spatial Technologies. Proceedings of the NATO Advanced Workshop, Ravello, Italy, 1-2 October 1999.* Amsterdam: IOS Press, 147–156.

Tilley, C.Y., 1994. *A Phenomenology of Landscape: Places, Paths and Monuments* 1ST edition. Oxford, Providence: Berg 3PL.

Tobler, W., 1993. *Three Presentations on Geographical Analysis and Modeling: Non-Isotropic Geographic Modeling; Speculations on the Geometry of Geography; and Global Spatial Analysis*. Santa Barbara: National Center for Geographic Information and Analysis.

Van Der Leeuw, S., 1981, information flows, flow structures and the explanation of change in human institutions, in S. Van Der Leeuw (ed), *Archaeological approaches to the study of complexity*. Amsterdam: Universiteit van Amsterdam (Cingula 6), 230–239.

Van der Leeuw, S.E., F. Favory and J.L. Fiches(eds), 2003. *Archéologie et systèmes socio-environnementaux: études multiscalaires sur la vallée du Rhône dans le programme ARCHAEOMEDES*. Paris: Editions du CNRS.

Van Ossel, P., 1992. *Etablissements ruraux de l'Antiquité tardive dans le nord de la Gaule*. Paris: Editions du Centre National de la Recherche Scientifique.

Van Ossel, P., 2010. De la 'villa' au village : les prémices d'une mutation, in J. M. Yante and A. M. Bulbo-Verleysen (eds), *Autour du 'village'. Établissements humains, finages et communautés rurales entre Seine et Rhin (IVe-XIIIe siècles) Actes du colloque international de Louvain-la-Neuve, 16-17 mai 2003*. Turnhout: Brepols Publishers, 219–236.

Van Ossel, P. and P. Ouzoulias, 2000. Rural settlement economy in Northern Gaul in the Late Empire: an overview and assessment. *Journal of Roman Archaeology* 13, 133–160.

Verhagen, J., M. van Leusen, B. Ducke, A. Millard and H. Kamermans, 2011. The bumpy road to incorporating uncertainty in predictive modelling, in J. Erzsébet, F. Redő and V. Szeverényi (eds), *On the Road to Reconstructing the Past. Proceedings of the 36th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, CAA 2008. Budapest, April 2-6.* Budapest: Archaeolingua, 580–587.

Verhagen, P., H. Kamermans, M. van Leusen, J.H.C. Deeben, D.P. Hallewas and P. Zoetbrood, 2007. First thoughts on the incorporation of cultural variables into predictive modelling, in F. Nicolucci and S. Hermon (eds), *Beyond the Artifact. Digital Interpretation of the Past. Proceedings of CAA2004, Prato 13–17 April 2004*. Budapest: Archaeolingua, 307–311.

Verhagen, P., 2007. *Case Studies in Archaeological Predictive Modelling*. Leiden: Leiden University Press.

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*.Archaeological Studies Leiden University, 71–121. Leiden: Leiden University Press.

Verhagen, P. and K. Jeneson, 2012. A Roman Puzzle. Trying to find the via Belgica with GIS, in A. Chrysanthi, P. Murrieta-Flores and C. Papadopoulos (eds), *Thinking Beyond the Tool. Archaeological Computing and the Interpretive Process. Proceedings of the 32nd annual meeting of the Theoretical Archaeology Group, 17-19 December 2010, Bristol*. Oxford: Archaeopress (BAR Archaeological Reports 2344), 123–130.

Verhagen, P., J. Joyce and M. Groenhuizen, 2016. Modelling the Dynamics of Demography in the Dutch Roman Limes Zone, in R. Hermans and A. de Rooy (eds), *3rd International Landscape Archaeology Conference 2014 Rome, 17th – 20th of September 2014*, Amsterdam: VU E-Publishing, 1-13.

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.* Leiden: Leiden University Press. (Archaeological Studies Leiden University 17), 19–25.

Verhagen, P., H. Kamermans, M. van Leusen and B. Ducke, 2010. New developments in archaeological predictive modelling, in T. Bloemers, H. Kars, A. van der Valk and M. Wijnen (eds), *The Cultural Landscape and Heritage Paradox: Protection and Development of the Dutch Archaeological-historical Landscape and Its European Dimension*. Amsterdam: Amsterdam University Press, 431–444.

Verhagen, P., H. Kamermans, M. van Leusen and B. Ducke, 2010. New developments in archaeological predictive modelling, in T. Bloemers, H. Kars, A. van der Valk and M. Wijnen (eds), *The Cultural Landscape and Heritage Paradox: Protection and Development of the Dutch Archaeological-historical Landscape and Its European Dimension*. Amsterdam: Amsterdam University Press (Landscape and Heritage Series 3), 431-444.

Verhagen, P., L. Nuninger, F. Bertoncello and A.C. Barba, 2015. Estimating the 'memory of landscape' to predict changes in archaeological settlement patterns, in S. Campana, R. Scopigno, G. Carpentiero and M. Cirillo (eds), *Keep the Revolution Going. 43rd Computer Applications and Quantitative Methods in Archaeology Conference (CAA 2015 SIENA), Mar 2015, Siena, Italy.* Oxford: Archaeopress, 623–636.

Verhagen, P., L. Nuninger, F. Bertoncello, F.P. Tourneux and K. Jeneson, 2013. Introducing the human factor in predictive modelling: a work in progress, in E. Graeme, T. Sly, A. Chrysanthi, P. Murrieta-Flores, C. Papadopoulos, I. Romanowska and D.W. Wheatley (eds), *Archaeology in the Digital Era. Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, 26-29 March 2012*. Amsterdam: Amsterdam University Press, 379–388.

Verhagen, P., S. Polla and I. Frommer, 2014. Finding Byzantine junctions with Steiner trees, in S. Polla and P. Verhagen (eds), *Computational approaches to the study of movement in archaeology: theory, practice and interpretation of factors and effects of long term landscape formation and transformation. Workshop of the Excellence Cluster Topoi, January 6 2011, Berlin*. Berlin, Boston: De Gruyter (Topoi, Berlin studies of the ancient world 23), 73–97.

Verhagen, P., I. Vossen, M. Groenhuijzen and J. Joyce, 2016. Now you see them, now you don't: Defining and using a flexible chronology of sites for spatial analysis of Roman settlement in the Dutch river area. *Journal of Archaeological Science: Reports* 10, 309–321.

Verhagen, P. and T.G. Whitley, 2012. Integrating Archaeological Theory and Predictive Modeling: a Live Report from the Scene. *Journal of Archaeological Method and Theory* 19(1), 49–100.

Verhart, L.B.M. and M. Wansleeben, 1992. The Meuse Valley Project: GIS and site location statistics. *Analecta* 25, 10.

Vermeersch, D., 1981. Le site archéologique du Marais de Famechon (Somme). Bilan provisoire. *Cahiers archéologiques de Picardie* 8(1), 147–156.

Warren, R.E., 1990. Predictive modelling of archaeological site location: a case study in the midwest, in K. M. Allen, S. W. Green and E. B. W. Zubrow (eds), *Interpreting space: GIS and archaeology*. London: Taylor and Francis (Applications of geographic information systems), 201–215.

Weiss, A., 2001. Topographic position and landforms analysis, in *Poster presentation, ESRI user conference, San Diego, CA*.

Wheatley, D., 1993. Going over old ground: GIS, archaeological theory and the act of perception, in J. Andresen, T. Madsen and I. Scollar (eds), *Computing the Past. Computer Applications and Quantitative Methods in Archaeology (CAA92)*. Aarhus: Aarhus University Press, 133–138.

Wheatley, D., 2004. Making space for an archaeology of place. *Internet archaeology* 15, 1–22.

Whitley, T.G., 2000. *Dynamical systems modeling in archaeology: A GIS approach to site selection processes in the greater Yellowstone region*. Pittsburgh: University of Pittsburgh.

Whitley, T.G., 2001. Using GIS to Model Potential Site Areas at the Charleston Naval Weapons Station, South Carolina: An Alternate Approach to Inferential Predictive Modeling, in Northern Illinois University and Argonne National Laboratory (eds), *GIS and Archaeological Predictive Modeling: Large-Scale Approaches to Establish a Baseline for Site Location Models*. Atlanta, Charleston, Raleigh: Brockington and Associates, Inc, 1–40.

Whitley, T.G., 2002. Modeling Archaeological and Historical Cognitive Landscapes in the Greater Yellowstone Region (Wyoming, Montana and Idaho, USA) Using Geographic Information Systems, in G. Burenhult and J. Arvidsson (eds), *Archaeological Informatics: Pushing The Envelope. CAA2001. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 29th Conference, Gotland, April 2001*. Oxford: Archaeopress (BAR international series 1016), 139–148.

Whitley, T.G., 2004a. Causality and cross-purposes in archaeological predictive modeling, in K. Fischer-Ausserer, W. Börner, M. Goriany and L. Karlhuber-Vöckl (eds), *[Enter the Past] The E-way into the Four Dimensions of Cultural Heritage. CAA 2003 I Computer Applications and Quantitative Methods in Archaeology I Proceedings of the 31st Conference, Vienna, Austria, April 2003*. Oxford: Archaeopress (BAR international series 1227), 236–239.

Whitley, T.G., 2004b. Re-thinking Accuracy and Precision in Predictive Modeling, in F. Nicolucci and S. Hermon (eds), *Beyond the artefact. Digital Interpretation of the Past. Proceedings of CAA2004. Prato 13-17 April 2004*. Budapest: Archaeolingua, 13–17.

Whitley, T.G., 2004c. Spatial Variables as Proxies for Modeling Cognition and Decision-Making in Archaeological Settings: A Theoretical Perspective. *Internet Archaeology* 16, 1-16.

Whitley, T.G., 2005. A brief outline of causality-based cognitive archaeological probabilistic modeling, in M. van Leusen and H. Kamermans (eds), *Predictive modelling for archaeological heritage management: a research agenda*. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek (Nederlandse Archeologische Rapporten 29), 123–138.

Wightman, E.M., 1975. The pattern of rural settlement in Roman Gaul, in H. Temporini (ed), *Politische Geschichte (Provinzen un Randvölker Gallien [Forts.], Germanien).* Berlin, New-York: Walter de Gruyter and Co. (Aufstieg und Niedergang der Römischen Welt 2(4)), 584–657.

Wightman, E.M., 1985. *Gallia Belgica*. Berkeley: University of California Press.

Figures

Tables

Abstract

The rural settlements of the Roman Somme (Northern France) are poorly understood in terms of site location. Although the physical landscape of the area is rather smooth, local variations influence the distribution of sites. Furthermore, the socio-economic context around Roman farms plays an important part in human behaviours of settlement creation. Predictive modelling constitutes an effective tool for dissecting settlement patterns and understanding their locational parameters through the quantification and evaluation of formal hypotheses. A specific methodology was tailored for the subject and inspired by theory-driven and cognitive predictive modelling approaches. It involves the creation of multivariate models through weighted map algebra, which are then confronted with the distribution of archaeological settlements in four micro-regions along the Somme River.

The correlation of the variables with archaeological location indicates that slopes, landforms and the relative distance to rivers are the main influential factors of the physical environment. Socio-economic parameters such as the relative distance to cities, secondary agglomerations and Roman roads are even more influential. Notwithstanding the lack of representation of settlements in the Late Roman period, site location follows similar trends from the 1st century AD to the end of the 4th century AD. Villas prefer economically well connected locations, as do stone-built and post-built settlements. Nevertheless, no parameter can be considered as deterministic in site location. This demonstrates the diversity of choices and influences which favoured the creation of Roman sites in the landscape.

Résumé

La compréhension de la distribution des établissements ruraux de la Somme romaine est particulièrement lacunaire. Si l’environnement physique de ce département est plutôt doux, les variations locales ont une influence sur l’emplacement des sites. De plus, le contexte socio-économique qui encadre les fermes romaines joue un rôle important dans les comportements humains entrant en jeu dans la fondation des établissements. La modélisation prédictive constitue un outil performant afin d’analyser la distribution des sites et en comprendre les paramètres d’influence, à travers la quantification et l’évaluation d’hypothèses formelles. Ce mémoire applique une méthodologie spécifiquement conçue pour ce sujet et qui est associée à la modélisation prédictive fondée sur la théorisation des comportements humains dans une approche explicative. Ceci implique la réalisation de modèles multivariés pondérés, lesquels sont ensuite confrontés à la distribution des sites archéologiques dans quatre microrégions longeant la rivière de la Somme.

La corrélation des variables avec la distribution des sites indique que les pentes, les formes du relief et la distance relative aux rivières constituent les principaux facteurs d’influence de l’environnement physique. Toutefois, les paramètres socio-économiques comme la distance relative aux cités, aux agglomérations secondaires et aux voies romaines sont encore plus influents. Malgré le manque de représentativité des établissements ruraux de l’Antiquité tardive, les mêmes tendances sont perceptibles dans la distribution des sites depuis le 1er siècle apr. J.-C. jusqu’à la fin du 4e siècle apr. J.-C. Les villas sont plutôt localisées dans des emplacements bien connectés au réseau économique, mais il en va de même pour les établissements maçonnés et en matériaux périssables. Néanmoins, aucun paramètre ne peut être considéré comme déterminant dans la distribution des sites. Cela évoque une importante diversité des choix et des influences favorisant certains emplacements du paysage.