Universität Augsburg
Fakultät für Angewandte Informatik
Institut für Geographie

Universität Augsburg
Fakultät für Angewandte Informatik

Developing a Data Driven Approach for the Analysis of Functional Settlement Pattern Considering Environmental Space

Inauguraldissertation zur Erlangung des Doktorgrades
vorgelegt von

Carolin von Groote-Bidlingmaier

1. Gutachter: Prof. Dr. Sabine Timpf
2. Gutachter: Prof. Dr. Jukka Krisp
3. Gutachter: Prof. Dr. Klaus Hilbert

Datum der Verteidigung: 22.02.2017

Abstract

It is difficult to draw reliable conclusions about prehistoric cultures due to inherently vague archaeological data. The lack of suitable methods for the quantitative analysis of settlement patterns in terms of uncertain as well as incomplete data is the motivation for this research. The aim is to gain the maximum of information about former settlements with a minimum of previous knowledge. To achieve that goal, the environmental surroundings in combination with the location are considered in the analysis. This is based on mainly two assumptions, namely that locations of settlements are influenced by environmental conditions and that settlements have different functions within the settlement-network.

Literature research was necessary in order to collect as many excavation site locations as possible. The result is the largest published collection of former settlement in the Amazon region. A conceptual data model is developed which fits the requirements of the available data and is implemented in a central database on a server in order to provide the data. This reduces redundancies and provides external access over the internet for all interested researchers. Consequently it facilitates the analysis of intra cultural settlement patterns. Additionally environmental variables need to be defined which are assumed to be potentially influencing.

A knowledge discovery process is developed which allows to further analyse the data. A Maximum Entropy Model is performed to see whether an environmental variable is influencing the outcome of the model. The variables with an ascertainable contribution are used for further analysis. The settlement type is determined using a clustering approach on the basis of the remaining environmental variables. To avoid that distant variables distort the cluster result, only environmental variables near the excavation site are considered. The definition of nearness is made using a rough boundary which is individually set for each parameter. The maximum nearness value is randomly selected and used as input for the cluster analysis. Various cluster runs with changing maximum nearness values are performed and compared using the consensus clustering approach. An optimal cluster solution as well as a consensus value are returned as a result which is used in order to calculate settlement function related suitability surfaces. These cost surfaces serve as basis for the

concluding territory analysis.

The developed methodology allows to derive scenarios of potential functional settlement patterns. The focus is on archaeological records of yet poorly explored cultures.

Zusammenfassung

Rückschlüsse über das Siedlungsmuster prähistorischer Kulturen zu ziehen ist aufgrund der unsicheren archäologischen Daten schwer. Das Fehlen geeigneter Methoden zur quantitativen Analyse von Siedlungsmustern im Kontext von unsicheren und unvollständigen Datensätzen ist die Motivation für diese Forschungsarbeit. Das Ziel ist es, ein Maximum an neuen Informationen über frühere Siedlungen bei minimalem Vorwissen zu gewinnen. Um das zu erreichen, werden die umgebenden Umwelteigenschaften in Kombination mit dem Siedlungsort analysiert. Dabei werden zwei Annahmen zugrunde gelegt, nämlich erstens, dass die Siedlungsorte durch die umgebende Umwelt beeinflusst werden, und zweitens, dass Siedlungen innerhalb des Siedlungsnetzwerks unterschiedliche Funktionen einnehmen können.

Eine Literaturrechere war nötig, um so viele Ausgrabungsorte wie möglich zu lokalisieren. Das Ergebnis ist die größte publizierte Sammlung von früheren Siedlungsorten im Amazonasgebiet. Dafür wurde in konzeptionelles Datenmodell entwickelt, welches den Bedingungen der verfügbaren Daten entspricht, und in eine zentrale, auf einem Server betriebene Datenbank implementiert wurde. Das reduziert redundante Datenhaltung und ermöglicht den externen Zugriff für interessierte Wissenschaftler über das Internet. Dadurch wird die Analyse intrakultureller Siedlungsmuster vereinfacht. Zusätzlich müssen die als einflussreich angenommenen Umwelteigenschaften definiert werden.

Es ist ein Prozessablauf entwickelt worden (Knowledge Discovery Process), der die weitere Analyse der Daten ermöglicht. Ein Maximum Entropy Modell wird angewendet, welches die einflussnehmenden Variablen für die Modellausgabe identifiziert. Die Variablen, die einen feststellbaren Einfluss haben, werden für die weitere Analyse verwendet. Der Siedlungstyp wird durch eine Clusteranalyse auf Basis der übrig gebliebenen Umwelteigenschaften bestimmt. Um zu vermeiden, dass weit entfernte Variablen das Ergebnis der Clusteranalyse beeinflussen, werden nur Umwelteigenschaften im nähreren Umkreis berücksichtigt. Die Definition der Nähe erfolgt mit Hilfe einer ungenauen, über Maximum- und Minimumwerte beschriebene Grenze, welche für jeden Parameter individuell festgelegt werden kann. Für eine Clusteranalyse wird ein Zufallswert maximaler Nähe innerhalb des Grenzinter-

valls bestimmt. Eine Vielzahl von Clusteranalysen mit wechselnden Zufallswerten wird durchgeführt und mithilfe eines Consensus Clustering Verfahrens verglichen. Als Ergebnis wird eine optimale Clusterlösung sowie ein Konsenswert ausgegeben, welche für die Berechnung der siedlungsfunktionsbezogenen Eignungsoberflächen verwendet werden. Diese Kostenoberflächen fungieren als Basis für die abschließende Analyse der Territorien.

Die entwickelte Methodik ermöglicht es, Szenarien über potentielle funktionale Siedlungsmuster zu entwickeln. Der Fokus liegt auf archäologischen Aufzeichnungen von bisher kaum erforschten Kulturen.

Table of Contents

List of Figures

List of Tables

Nomenclature

AIC	Akaike Information Criterion
ANN	Artificial Neural Network
ASC	ESRI ASCII Raster Format
AUC	Area Under the Curve
BIC	Bayesian Information Criterion
CSR	Complete Spatial Randomness
CSV	Comma Separated Values
DEM	Digital Elevation Model
DIC	Deviance Information Criterion
GDAL	Geospatial Data Abstraction Library
GIS	Geographic Information System
GKD	Geographical Knowledge Discovery
HPP	Homogeneous Poisson Process
IBGE	Instituto Brasileiro de Geografia e Estatística
IPHAN	Instituto do Patrimônio Histórico e Artístico Nacional
KDD	Knowledge Discovery from Databases
MCE	Multi-Criteria Evaluation
MEM	Maximum Entropy Model
NF	Normal Form
OGC	Open Geospatial Consortium
PCA	Principle Component Analysis

PRONAPA Programa Nacional de Pesquisas Arqueológicas

PRONAPABA Programa Nacional de Pesquisas Arqueológicas na Bacia Amazônica

ROC Area Under the Receiver Operating Characteristic

SRTM Shuttle Radar Topography Mission

1 Introduction

When humans settle in a specific area they interact with the environment surrounding them. Settlement decisions in prehistory were determined by the social situation (e.g. forced by conflict) or environmental criteria in order to fit the peoples' needs and capabilities. Thus, the location is a compromise which considers all influences that may affect a population (Sollars 2004). A major issue was the availability of food which is limited to the goods which were accessible. The way people interact with the environment varies according to the availability of necessary goods and the perceived capabilities of the surrounding space.

> The lifestyle of any given community is closely tied to environmental considerations. (Sollars 2004, p. 258)

Only the findings of excavation sites provide evidence about some of the materials and goods used and thus give a rough picture of how people perceived and used the environmental surroundings. However, archaeological data is inherently vague. On the one hand the number of known excavation sites might not – and probably does not – correspond to the actual number of former settlements. On the other hand the excavated findings do not represent a complete range of formerly used goods and materials. Reasons can be the climate conditions, the soil condition, and other environmental circumstances such as flooding, as well as the applied archaeological practice (Funari 1995). Considering the environmental surroundings – including the topography, as well as the accessibility of water supplies or exploitable resources – extends the available findings related data.

> In the Amazon basin [...] only the comparison of cultural sequences from a variety of ecological settings can provide new insights for testing competing models of cultural developments. (Barreto 1998, p. 579).

In terms of Brazilian archaeology political and historical circumstances had an influence on the archaeological fieldwork. During the military invention in 1964 and the subsequent military government

> Brazilians could no longer profess a different view without being considered as external enemies. (Rodrigues et al. 1984, p. 226)

These circumstances led to a predominant archaeological alignment which focused on descriptive and classificatory practices rather than interpreting the past (Barreto 1998, Funari 1995). Despite an ongoing change towards more problem-oriented research in Brazilian archaeology since the 1980's there is still a lack of standardized research methods and knowledge (Neves 2008, Barreto 1998). That also led to confusing and research depending definitions and nomenclatures of archaeological cultural divisions and subdivisions. In Amazonian Archaeology a common categorization of cultures is based on differences in ceramic styles (see chapter 2.1.1). It would be of interest to archaeological research

> ... if data on ceramic chronology could be matched by more information on things such as settlement patterns, site occupation chronology, intra-site spatial patterns and so forth, but this is not available in the vast majority of cases. (Schaan 2007 (as cited in Neves 2008, p. 365))

Another settlement pattern related aspect is the existence of functional settlement types. Settlements differ in size, environmental surroundings, time of occupancy, etc. which suggest functional related settlement types.

> Also, groupings of sites into fragile categories of phases and traditions often hid other, more informative site classifications, such as site function. (Barreto 1998, p. 577)

The available information about the former settlements varies considerably. For some excavation sites only the location and, based on the ceramic findings, the former culture(s) are known. For other excavation sites the arrangement of buildings as well as the intra site relationships are known. In order to use as many excavation sites as possible the settlement patterns should be analysed regardless of the available information. Therefore, the aim of this work is to *develop a method in order to gain as much information about the functional settlement patterns as possible using only the location of the excavation site and its surrounding environment*. This allows the identification of functional settlement types without additional knowledge as well as an assessment about the importance and range of each environmental parameter.

The developed approach is referred to as functional settlement pattern analysis and is based on the following assumptions.

- The settlements of one observed group (e.g. a specific culture) are seen as a system of interacting settlements which implies an exchange of goods. Additionally, the culture – and thus all settlements – benefits from the functions (e.g. defence against enemies).

- Settlements provide different functionalities, and the suitability of a location depends on the function. This also means that some of the functionalities rely on the environmental surroundings. E.g. trading functionalities can only be close to trading routes.

- Certain environmental surroundings are needed in order to serve as suitable locations. This includes goods required for survival such as drinking water and food. Whereas the environmental conditions may vary according to the capabilities and needs of the settling group. It makes a difference whether a settling culture is able to grow its own food or not.

- The distance to important resources may not exceed a maximum distance but may vary according to the type of resource. This means that the distance which a culture is willing to overcome varies according to the resource.

- The relevant resources may vary according to the culture and functionality. This is due to the varying capability and needs of a culture as well as the function related specific requirements.

- The location of the settlement allows to draw conclusions about the settlement type. Thus, the specific combination of environmental parameters is important for the settlement function.

- The suitability of a location can be assessed on the basis of the environmental surroundings. Thus, a minimum number of environmental parameters is needed in order to be considered as suitable location.

The introduction given above leads to the following research questions:

1. How can the locations of the excavation sites be stored to function as basis for further analysis?

2. How can environmental information be considered in order to identify specific functional settlement patterns? How can the influence of the environmental variables be determined?

3. How can knowledge about functional settlement patterns be gained? How must this knowledge discovery process be designed in order to consider the various settlement and subsistence strategies?

Based on these questions the following research objectives are derived:

1. A data model shall be created which can store the locations and other available information of the former settlements.

2. A model needs to be developed which fits characteristics of Amazonian Archaeology – namely the inherently vague data – and the specific environmental conditions.

3. A knowledge discovery process shall be developed which allows the identification, characterisation and analysis of functional settlement pattern. The developed approach is dynamic to be applicable to all cultures regardless of the cultures capabilities and needs. Additionally, the method needs to work with an arbitrary number of archaeological subdivisions of the input locations (e.g. based on ceramic styles, language groups etc.). Statistical methods should be applied to underline the relevance (or irrelevance) of the surrounding resources. The results should be reproducible, thus subjective input need to be minimized.

4. This work shall provide a method which allows archaeologists to develop theories on settlement patterns of the observed culture.

The approach focusses on developing a method which allows functional settlement pattern analysis despite the variety in data quality.

In order to analyse settlement patterns the location of the excavation site needs to be known. The sources of archaeological data in the Amazon Basin are manifold. Besides a high number of publications with maps (some including a small description of the location), the Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN) provides a database with the known excavation sites (link to the database: `http://portal.iphan.gov.br/sgpa/?consulta=cnsa`). Unfortunately the database does not contain any coordinates which makes it useless for any kind of spatial analysis (see A for an excerpt of the database). Thus, reliable data about the former settlements needs to be collected in order to analyse the functional settlement pattern of pre-colonial cultures in the Amazon Basin. This implies information about the location as well as the environmental surroundings. Most of the excavated settlements are published as paper maps with a brief description of the location, thus exact coordinates are not known. Therefore, the location of the excavation sites needs to be identified in a first step.

As soon as the locations are known, the settlements can be analysed. The spatial distribution of the locations is analysed to make sure that neither a regularly dispersed nor a random pattern of the location occurs. If a clustered pattern can be identified, further distinctions into functional patterns can be made. Excavation sites can be subdivided into several groups (or hierarchies) based on various criteria. This allows the consideration of site size (which is used as an estimation for the former population size) (e.g. (Drennan and Peterson 2004, Johnson 1977, Blanton 1976)) or other functionalities (e.g. residence of the chieftain) (e.g. (Steponaitis 1981, Peebles and Kus 1977)). This requires knowledge or an expert opinion about the settlement hierarchy or social function in advance. But what if there is no such information and subjective interpretations should be avoided. For the majority of the excavation sites in the Amazon only the location and the culture (based on the ceramic styles) are known. Only few excavation sites have information about the size. In that case no distinction can be made in advance. An alternative would be to use a data driven, explorative approach in order to derive more information about the location. This means that the functions are derived based on the environmental surrounding rather than on findings or the site itself. Consequently, the environmental parameters as well as the potential importance of those need to be determined. These results are used to find agglomeration patterns in order to identify the settlement type. The environmental variables in combination with the determined settlement function can then be used for the calculation of the suitability surface for a specific location. Based on the analysis, the territory with regard to the cultural or functional preferences can be determined. Due to the data driven approach, the developed method can be applied to every archaeological record where the location is known. Neither additional input data – besides the environmental information – nor interpretations of the data is needed.

The developed approach aims to facilitate settlement pattern analysis based on archaeological excavation sites. It focuses on locations and their environmental surroundings rather than expert knowledge. Based on the applied model (e.g. consideration of specific cultural characteristics) the presented approach can therefore be used to verify archaeological hypotheses. Additionally, the resulting suitability surface may also lead to new hypothesis and might help to identify further former settlements.

To summarize the above, only very little is known about the settlement patterns of pre-colonial cultures in the Amazon. The aim is to develop a procedure which gains knowledge despite the fact that only little information is available. This is done by considering the environmental surrounding of an excavation site. The presented approach shows which steps are necessary in order to achieve this goal. However, the selection of environmental parameters (as well as the classification of these) may not be sufficient with regards to the analysed culture. Thus this research does not aim for a comprehensive functional settlement pattern analysis but rather for the presentation of the developed process.

Chapter 2 provides an overview of archaeological research in the Amazon Basin as well as settlement pattern related studies. The third chapter describes the developed knowledge discovery process. This is applied in a case study using the Konduri and Guarita culture (grouped by ceramic style) which is presented in chapter 4. The results are discussed in chapter 5. Chapter 6 provides a conclusion about the developed approach including the major findings as well as future work.

2 Theoretical Background

Due to the interdisciplinary nature of this research several aspects are covered in this dissertation. An overview of the archaeology in the Amazon is given in chapter 2.1. This chapter is subdivided in two sub-chapters. The first focusses on the chronological framework (chapter 2.1.1) whereas the second is about the theories of settlement history (chapter 2.1.2). The second chapter covers the relevant aspects of spatial analysis in the archaeological context. This covers research on settlement pattern analysis (chapter 2.2.1) which is followed by the sub-chapter on site catchment and territory analysis (chapter 2.2.2). The aspects of predictive modelling are described in the last chapter (chapter 2.2.3).

2.1 Archaeology in the Amazon

Providing a complete review of archaeological research in the Amazon lies beyond the scope of the presented research. This chapter gives a broad overview of the pre-colonial cultures in the Amazon. The Amazon Region is a large area covering a high percentage of northern South America and therefore has a huge variety in geographical (see chapter 4) as well as archaeological aspects. In contrast to the cultures of higher complexity (e.g. Inka or Maya in the Andes and Middle America) there is only very little information about the peoples in Amazonia (Hilbert 1977), due to multiple reasons. As Barreto (1998) stated by the example of Brazilian archaeology, the size of the country, the lack of resources and governmental support and the difficulties due to the climate conditions have influenced the archaeological work. The humidity in the Amazon Region has decreased the number of materials which survived the past centuries and can still be found today. Items made of wood, feathers, leather or other organic material are more likely to be found in arid regions than in the Amazon Basin (Hilbert 1977). Furthermore, erosion and floodings can lead to site destruction (Denevan 1996).

The number of archaeological surveys increased in the 1970s (Barreto 1998). That is the reason why the Brazilian archaeology, in contrast to the Northern American archaeology, remained in what Willey and Sabloff (1993) described as the classificatory-descriptive stage. The principle goals were classification and description of excavation sites without any interpretation or explanations (Lewis 1996). A change in practice was initiated by the work of Clifford Evans and Betty Meggers who defined cultural sequences by dating and analysing ceramics. A group of archaeologist from Brazil and North America – organized and administered by Meggers and Evans 1965 - 1971 – have joined forces and initiated a program named PRONAPA (Programa Nacional de Pesquisas Arqueológicas). Their main goal

> was the understanding of the processes by which successive groups of pre-European immigrants with different subsistence patterns adapted to the diverse environmental conditions within Brazil (PRONAPA 1970, p.1).

That also led to a shift of focus which turned from pre-ceramic sites to ceramic occupations (Schmidt Dias 1995 (as cited in Barreto 1998)). Evans and Meggers (1973) stated that the Latin American archaeology

> has undergone a tremendous expansion [...] during the past two decades (Evans and Meggers 1973, p. 257)

with a growing number of fieldworks and publications. Whereas the focus was on south and central Brazil in the beginning the Amazon was focus of the PRONAPABA (Programa Nacional de Pesquisas Arqueológicas na Bacia Amazônica) project Amazonian research now has a problem of interpreting the patterns rather than suffering from a lack of material (Eriksen 2011).

2.1.1 Chronological Framework

With the increasing number of ceramic sites it was possible to develop a chronological framework for the findings (Hilbert 1977) which was necessary to achieve the goal of PRONAPA (1970), namely defining cultural sequences by dating and analysing ceramics.

In his multi-volume works Steward (1947) divided the South American tribes into four cultural regions:

- the Marginal tribes,
- the Andean civilizations,
- the Tropical Forest tribes and
- the Circum-Caribbean tribes.

Evidence can be found for all but the Andean civilizations in the Amazon region although the majority of findings are related to the tropical forest tribes (Hilbert 1977). The Marginal tribes are described as hunters and gatherers with neither agriculture nor pottery. Arrowheads are usually made out of wood or bones but were lost due to humid climate conditions, only some stray findings made of stone exist. The Tropical Forest tribes as well as the Circum-Caribbean tribes had knowledge of ceramic manufacturing. In volume three of Steward's Handbook of South American Indians *The Tropical Forest Tribes* the theoretical framework of cultural ecology was introduced. The attempt was to use the adaptions to the local ecology to explain the social and economic organization (Eriksen 2011). Neves (1998b) mentioned that the Tropical Forest tribes

> were characterized by a curious blend of adaptive traits (Neves 1998b, p. 625)

which are difficult to apply. Additionally, it reduces a complex culture to environmental variables (Eriksen 2011). The Tropical Forest tribe is typically described as subsistence farming units with simple technology and with no ability to produce food surpluses - and therefore economic surpluses. Mainly negative terms were used to describe their behaviour in order to distinguish the Tropical Forest Tribes from the Circum-Caribbean tribes (Lathrap 1970). Due to these limitations they were unable to improve their social structure, political centralization and craft specialization (Viveiros de Castro 1996) and no cultural complexity was possible (Lathrap 1970). The tribes of the Circum-Caribbean group are distinguished among other things by having an influential chiefdom with large communities of about 1000 or more members, developed patterns of warfare, a religious system including adoration of deities and a professional priesthood (Lathrap 1970).

Findings at the excavation sites are used to further divide the tribes into subgroups based on pottery styles (Heckenberger and Neves 2009, Hilbert 1968, Meggers and Evans 1961). The primarily used cultural chronology distinguishes between four horizons (Neves 2008). Horizon refers to a historical unit which links together contemporaneous cultural criteria dependent on the combination of broad spatial and short temporal dimensions (Willey and Phillips 1958, Willey 1945). It represents a relative chronology of the former peoples and is based on two fundamental aspects. One is the similarity among style groups in each horizon (in the Amazon mainly ceramics), the other is their relative position in the time sequence to facilitate the temporal allocation of more local styles (Kroeber 1944). Willey and Phillips (1958) distinguish between horizon style and horizon whereas a horizon style is a narrower concept of the latter. The authors roughly define a horizon style as

> the wide distribution of a recognizable art style. On the assumption of historical uniqueness of stylistic pattern, coupled with the further assumption that styles normally change with considerable rapidity, the temporal dimension is theoretically reduced to a point where the horizon style becomes useful in equating phases or larger units of culture in time that are widely separated in space. (Willey and Phillips 1958, p. 32).

In contrast to the definition of a horizon style concept, the horizon concept does not presume a level of aesthetic development but can be applied to any kind of archaeological evidence that indicates a rapid spread of new ideas over a wide geographic space (e.g. highly specialized craftwork types) (Willey and Phillips 1958). This enhancement facilitates the adaptation of the terminology to the Amazon Archaeology with the partly absence of artistic elements. As Hilbert (1977) mentioned, handcrafted ceramic products of Circum-Caribbean tribes are more diverse and abundant, as can be seen from the example of the Tapajó tribe which was assigned to the Incised-Punctated group. In contrast to that, ceramic findings of other tribes do not necessarily have to be that manifold, which can be seen in (Simões 1981) by the example of Mina-ceramics and can be the remains of objects of daily use. Mainly due to the efforts of PRONAPA and PRONAPABA a standardized terminology was implemented (Neves 2008, PRONAPA 1966). In PRONAPAs (1966) nomenclature the term horizon style is not listed but horizon is defined as an association of traditions which occur in the same temporal dimension and cover various geographical areas.

The term horizon was discarded in the Amazon Region and is now replaced by the term tradition (Schaan 2001, Roosevelt 1991b, Meggers and Evans 1983 (as cited in Neves 2008, p. 365)). One reason was the cultural historical assumption that all the horizons' origins were located outside the amazon. The other reason was the limited time span inherent with the concept whereas tradition was sometimes defined as counterpoise with a greater temporal dimension. Tradition is a broadly used word with the absence of a clear definition. Unlike the horizon (or horizon-style) Willey and Phillips (1958) define a tradition as a unit with a bigger temporal span which represent

> temporal continuity represented by persistent configurations in single technologies or other systems of related forms (Willey and Phillips 1958, p. 37).

This definition sees pottery development as a long-term process with recognizable techniques or styles. A more general definition was provided by McGregor (1950).

He considered a tradition to be some sort of human characteristics or behaviour which is passed on to the next generation and is mostly technologically orientated (e.g. pottery, house types, etc.). Goggin (1949) additionally considered environmental factors to identify culture-environment correlations. A tradition according to PRONAPA (1966) is defined as an element or technique which persists over a longer time period.

The traditions are further divided into phases and sometimes subphases which are considered to be equivalent to separate cultural units (Neves 2008). As a result of PRONAPABA an overview of the increasing number of excavation sites and thus cultural units were published by Simões (1972). The report includes some radiocarbon measurements as well as the assignment to traditions and phases and revealed the cultural richness of the Amazon Basin. While the term phase was taken into consideration by PRONAPA (1966), the term subphase is not listed. According to this report a specific phase is characterized by the similarity of artifacts, similar geographical or temporal positions but also residential habits or resemblance of chieftains. In contrast Willey and Phillips (1958) provide a definition that describes a phase as an archaeological unit which is distinct from all other units by possessing traits, spatially limited to a region or locality and a relatively short time span. Additionally they point out that the distinction between these units are independent of the related culture. This definition extended the definition provided by (Kidder et al. 1946) in terms of spatial limitations. Meggers (1987b) claims:

> When the National Archaeological Programme (PRONAPA) began, we believed that an archaeological phase was an abstraction without any ethnographic basis. Now, however, we believe that phases, defined in terms of sequential series, represent separate entities, while traditions, defined in terms of phases which share common features, represent tribal or linguistic entities. (Meggers 1987b, p. 13 as cited in Funari 1995).

Similar to a phase, culture is defined as a collection of culture traits which can be more or less distinguished from others and whose separate traits are applied by all or by a selected group of individuals (Taylor 1948). The terms phase and culture are sometimes used synonymously.

Initially, four traditions were identified based on the different pottery styles. The four traditions (former horizons) were the following:

- Zone-Hachured (approx. 2500 BP - 1500 BP),
- Incised Rim (approx. 1900 BP - 1200 BP),

- Polychrome (approx. 1400 BP - 700 BP) and
- Incised-Punctated (approx. 1000 BP - 500 BP)

(Meggers and Evans 1983, Hilbert 1977, 1968, Meggers and Evans 1961). In contrast to the subdivision presented by Steward (1947) an epistemological shift towards grouping based only on material grounds was initiated by Meggers and Evans (Almeida and Neves 2012). For a general idea of the classification procedure see also table B.01.

The Zone-Hachured tradition is characterized by the use of simple vessels with a certain scratching technique. The forms on the findings are framed by a broad line. The form itself is filled with thin parallel or crossed hatching. A broad rim of a vessel is a characteristic for the Incised Rim tradition. The horizontal or slightly inclining rims have carvings on them – some of them enhanced with red color. Both traditions are assigned to the tropical forest tribes (Hilbert 1977). The combination of polychrome paintings, incision and anthropomorphic modelling is typical for the Polychrome Tradition (Roosevelt 1980). Most of the related cultures are classified as Circum-Caribbean tribes. Besides precise carvings and dottings, ceramics of the Incised-Punctated tradition were often decorated with animal or human shaped symbols. This tradition is assigned to the tropical forest tribes, however some cultures such as the Tapajó are categorized as Circum-Caribbean tribes because of their high quality ceramics (Hilbert 1977). The dates given above are an approximation and may differ depending on the region due to time lag of the spread which may be subject to revision with a better knowledge of the chronology (Meggers and Evans 1961 and see also figure C.01). In addition to the four previously defined traditions, further (earlier) traditions need to be added to the chronology because they were not identified when the four traditions listed above were classified (Roosevelt 1995, Simões 1981). Simões (1972) assigned the known phases to the related tradition. Several phases do not match the criteria for the known traditions and are categorized as other traditions. However, some of these phases are labelled with either M or T meaning Mina tradition or Tupiguarani tradition (see table D.01).

In addition to the presented periodizations others were published by e.g. Lathrap (as cited in Heckenberger et al. 2008) who provided another terminology, Roosevelt (1997) who brings up a more general historical view (such as Paleo-Indian tropical foragers, shell-midden peoples etc.) or Heckenberger et al. (2008) who prefers a more temporal listing with e.g. archaic (pre-3000 BP), regional development and diversification (2000 BP - 1500 BP) or late prehistoric *classic* (1500 BP - 1000 BP).

The data used in this dissertation (see chapter 3.1 for a more detailed description of the data set and its creation) is a collection of excavation sites based on several archaeological publications. Due to the widely applied division into traditions (former horizon) and cultures (or phases) in these publications, this chronological framework is used to distinguish between the findings within this thesis. The term tradition is used to define the superordinate level of similar ceramic styles, which is subdivided into several cultures. The terms culture and phase are used synonymously, despite the fact that the meaning is not the same. That has no effect on the results because a culture or phase in this research is a term to label a subgroup of a specific tradition but with similar traits. No new assignment to a group or subgroup is undertaken but the given classifications of the data sources are used.

2.1.2 Theories of Settlement History

For more than 10.500 years people have been living in the Amazon (Roosevelt et al. 1996). While the first people were hunters and gatherers, the first agricultures existed about 6.000 (or more) years ago (Heckenberger and Neves 2009). The first chroniclers went to the Amazon about 500 years ago – Francisco de Orellana was the first who travelled from the source to the mouth of the Amazon river in 1541 – and still a lot of knowledge about the living habits of the societies in the Amazon is based on their observations (Neves 2008). In their reports one can find descriptions of large settlements, some of them extending over several miles and chieftain controlled huge territories including several settlements. Some of them were equipped for military purposes and some had beautiful pottery (Porro 1994, De Carvajal 1942).

The process of understanding the emergence of domestication, sedentism, and social stratification, thus reconstructing the Amazon history, caused and continues to keep alive a debate among researchers (Hornborg 2005). Regrading the Tropical Forest tribes, Lowie's (1948) opinion was that water was important and gave cultures the ability to spread their influence over a larger area. Based on this, a distinction between floodplain (so called várzea), and hinterland (so called terra firme) societies was made, which Neves (1998b) describes as a "sacred axiom" and has been characterized as The Standard Model by Viveiros de Castro (1996). Although Steward (1947) recognized different ecological circumstances such as different forest types, the tropical forest was seen as one uniform environment with unfriendly living conditions and these restricting factors limited the Amazonian population (Hames and Vickers 1983). The tropical rain forest is often seen as a symbol for the dominance of nature over humans and for the theory that life was determined by ecological factors (Whitehead 2010). The suggestion was, that the different subsistence, social and demographic patterns were related to the varying environmental surroundings and the ecological adaptation (Roosevelt 1980). These assumptions are supported by Meggers who described the Amazon as a counterfeit paradise (Meggers 1971). She hypothesized that a diffusion of ceramics of early cultures out of western or northwestern South America into the Amazon had taken place and in fact does not have its origin in the Amazon (Meggers 1979, 1987a). One example is the Marajoara phase, which is described as a complex society that emigrated to Marajó Island and decayed in the poor environment of the island (Meggers and Evans 1957). Based on the current state of research at that time it seems as if none of the known ceramic complexes had their origin in the Amazon. The Amazon was said to be a reservoir of different cultural influences rather than a center of origin of cultures. Whereas Zone-Hachured and Polychrom cultures appear to have their origin in the West, the other two horizons seem to have their origin in the North West of South America (Hilbert 1977). Meggers further identifies Holocene environmental changes – and their inferior environmental conditions – as a major limiting factor by linking knowledge gaps in the time line to arid intervals (Meggers 1988 as cited in Neves 1998b). That underlines Meggers's (1971) theory of the Amazon as being unsuitable for human life due to the lack of cultivable land (Roosevelt 1980). The theory is not uncontentious and some critics consider that it is based

> on a blend of monocausal environmental possibilism and diffusionism, modifying some perspectives along the way. Indeed, it has been such willingness to modify ideas that makes it hard to overlook the arguments [...] (Neves 1998b, p. 626).

Additionally, the cultural differences cannot be explained by the assumed equal environmental conditions (Roosevelt 1980).

It is worth mentioning, that PRONAPA and PRONAPABA are not without controversy. Funari (1995) criticized that this relatively small group of North American and Brazilian researchers controlled the excavation, funds, publications, posts, and spread of different or foreign perspectives and took advantage of their leading position. This accusation is supported by Roosevelt (1991a) who claimed that evidence which disproves the announced theory was often not allowed to be published or was ignored (as Neves 1998b stated). One example is Miller's (1992) research in Rondônia which leads to the assumption that there was a continuous human occupation in the area from 8320$\pm$100 BP. Funari (1995) notes that a non-positivist approach and archaeology as an experimental science was improper to Brazilian culture. Other authors state that PRONAPA did not attempt to explain the cultural development but that the major cause was the different research focus of PRONAPA (Barreto 1998). The Standard Model was dominant well into the 1990s although it has been criticized by several other researchers (cf. Roosevelt 1994, Balée 1993, Moran 1993, Carneiro 1961).

Another theory is based on the hypothesis that the alluvial floodplains of the tropical forest were a center of cultural development – the so called cardiac model – and therefore the oldest ceramics should have been found in the central Amazon floodplains (Lathrap 1974, 1973b,a). This theory is contrary to Willey's 1962, where the Amazon plays only a minor role according to the cultural historical influence. Lathrap (1973a) disagrees with the "erroneous" assumption of the tropical forest being a homogeneous environment and thus disabling long-distance trade networks as well as with the environmental determinism inherent with the *cultural ecology* concept (Eriksen 2011). The cultures were aware of the potential of the environment and had the capabilities to process materials, such as fibers, oils, woods, etc. and

> represent a total and non-destructive mastery of its environmental settings. (Lathrap 1973a, p. 171).

That also includes the ability to cultivate plants and house gardening (Heckenberger and Neves 2009, Neves 1998b). The confirmation for the presence of cultivated plants could not be found in the beginning (Lathrap 1970) but subsequent research provided evidence that several groups were using domesticated plants about 4.000 BP (Whitehead 2010, Heckenberger and Neves 2009, Balée 1994). In addition to that, research has shown the ability of producing – whether intentionally or not – and using very fertile soils (so called Amazonian Dark Earth or Terra Preta – see bellow) (Whitehead 2010). The importance of early agriculturalists and their dispersal is outlined in a worldwide review by Diamond and Bellwood (2003). Although the theories and proposed models about the shift from food procurement to food production systems vary, the importance of agriculture is undisputed (Oliver 2008).

Lathrap (1970) claims, that unlike e.g. art style or technology, the primary language is learned early and is not easily changed. Thus he provided a hypothesis of lowland distribution based on the two main linguistic stocks, namely the Arawakan and the Tupi-Guarani, and their common origin in the Amazon floodplain around 5000 BP (see also Neves 1998b). Both linguistic stocks are supposed to have developed from an "Ancient Amazonian Polychrome Tradition" but radiocarbon analysis for sherds in the central Amazon (Heckenberger 1998) and in the lower Negro basin (Neves 1998a (as cited in Neves 1998b)) resulted in younger dates, hence the hypothesis is still disputed (Neves 1998b). Consistent with Lathrap's (1974) theory on early pottery in the Amazon, Roosevelt (1995) dated the beginnings of ceramic production back to 8000 BP which makes it the earliest known example of pottery in America (Whitehead 2010). Evidence has been provided based on sherds from Tapérinha (a shellmound near Santarém) and Pedra Pintada (a cave-site near Monte Alegre). It seems that the findings differ from the Ancient Amazon Polychrome Tradition envisioned by Lathrap and his group (Neves 1998b). In contrast to Meggers' and Evans' (1957) view on the Marajoara phase (see above), Roosevelt (1991b) hypothesized that the polychrome ceramics were records of a locally developed chiefdom. Based on the work of early chroniclers and other archaeologists Roosevelt (1991b) expected the Marajoara and other Amazonian chiefdom to be complex cultural communities. Some of the described traits are the location along floodplains, expansionist warfare, intensive agriculture and exploration of aquatic fauna.

In all theories the environment and its limiting factors are given as main reason for the migration and emigration processes.

> In the end, it seems that, regardless of the differences among their cultural historical models, Roosevelt employs the same heuristic concepts – rooted in one way or another in some form of ecological determinism – as those previously employed by Meggers and Lathrap in their explanations (Neves 1998b, p. 630)

and also by (Carneiro 1995, Denevan 1976, Carneiro 1970) and (Denevan 1966) and others.

Another approach is taken by Denevan (1996) who does not distinguish between cultures from the floodplain and from the hinterland but rather proposes a combined use. Prehistoric roads and causeways in the Amazon Region underline the idea of relations between várzea and terra firme and indicate the presence of large and complex societies (Whitehead 2010, Woods and Glaser 2004). Denevan (1996) argues, that the floodplain is a high-risk habitat due to potential flooding and therefore not suitable as an autonomous settlement site. While Roosevelt (1980) located agriculture and dense population in the floodplain, Denevan (1996) proposes valley-side bluffs adjacent to the river channels (as shown in figure 2.1). Meggers (1984) believed that there were only few camps or fishing stations which were mainly temporary during low water. In Denevan's (1996) opinion, Meggers was essentially correct according to the relevance of the floodings but underestimated the importance of bluff zones and their agricultural potential and hence presents an integrated bluff/várzea strategy.

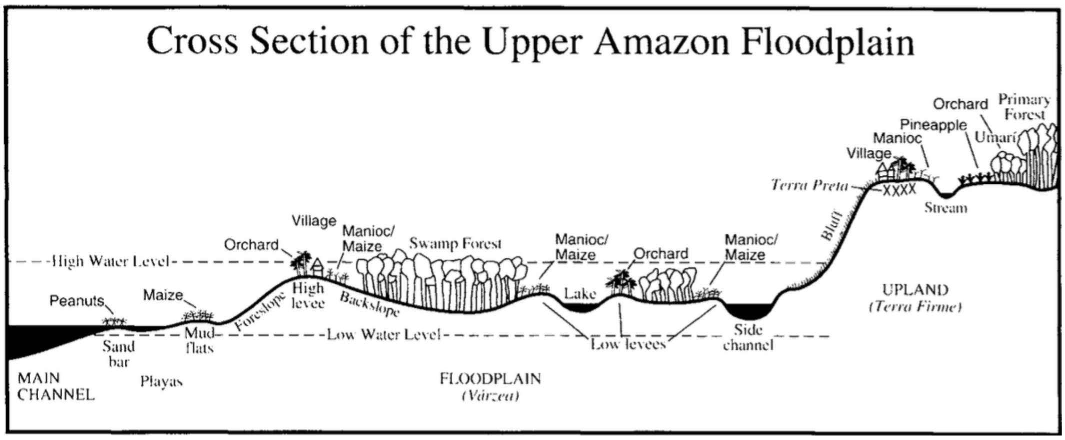


Figure 2.1: Representative cross section of the Upper Amazon floodplain near Iquitos showing the main channel, side channel, water levels, natural levees, bluffs, villages Terra Preta site (Denevan 1996)

With a growing emphasis on environmental diversity it became obvious that this division between terra firme and várzea is not sufficient due to the pedological, botanical and zoological variety (Woods and Glaser 2004, Viveiros de Castro 1996, Moran 1995, 1993, Prance and Lovejoy 1985). The diversity of environmental conditions, and therefore the agricultural expansion, cannot be covered by models using unified processes of site or trait diffusion. Complex and variable processes are needed instead (Heckenberger and Neves 2009). In Hornborg's (2005) opinion these transformations of material culture and language are the results of cultural development (Eriksen 2011). The debate on human adaptation to the Amazonian environment is still ongoing and new research techniques will lead to new insights. It seems that a shift from Amazonia as false or counterfeit paradise towards a complex environment with various human behaviours is taking place (Eriksen 2011, Whitehead 2010, Oliver 2008, Woods and Glaser 2004). This view is supported by the perspective of *Historical Ecology* with its basic premise that cultures in the Amazon did not adapt to nature but rather used human creativity, technology and engineering, and cultural institutions to create the world according to their ideas and needs (Oliver 2008, Denevan 2001, Balée 1998). The focus lies on the historical landscape which is described as a multidimensional entity with spatial and temporal characteristics. The landscape has been modified by humans and their activity so that human behaviour can be inferred from it.

> The landscape is like a text, but not one that is readily accessible to historians' and epigraphers' methods because it is not written in a decipherable script, but rather is inscribed in a subtle, physical sense by learned, patterned behavior and action – what anthropologists tradition-

ally refer to as *culture*. (Balée and Erickson 2006, p. 2)

In contrast to other human-environmental approaches such as natural selection, kin selection or climate change, historical ecology also includes prehistory – with a landscape as a result of previous occupants – and therefore is anthropocentric (Oliver 2008, Balée and Erickson 2006). Humans are seen as *keystone* species (Mann 2002) whose environmental interferences can lead to various diverse landscapes (Balée 1998, Botkin 1990, Connell 1978). The shift towards the historical ecology is also important in terms of the debate about Terra Preta and its genesis as most scholars classify Terra Preta as anthrosol (Eriksen 2011). The importance of house gardening and agriculture is emphasized in many archaeological publications (Whitehead 2010, Heckenberger and Neves 2009, Woods and Glaser 2004, Neves 1998b, Denevan 1996, Balée 1994). One of the main environmental resources for a productive agriculture is the soil. The soils in the Amazon Region were long seen as low in nutrients. Increasing emphasis on the environmental diversity and a shift from continental or regional depictions to microscale reveals a great variety in the soil landscape (Woods and Glaser 2004). Besides nutrient poor soils such as Oxisols, Ultisols, and Acrisols, fertile anthrosols are found (so called Indian Black Earth, Terra Preta (do índio or Amazonian Dark Earth) (Glaser et al. 2004). Katzer (1903) assumed that the origin of the Amazonian Dark Earths was cultural in nature – in contrast to the European Chernozems which developed naturally. Although the origin of this soil type is not yet clarified, it is known that the cultures – intentionally or unintentionally – created Terra Preta to increase the carrying capacity (Woods and Denevan 2009, Woods and McCann 1999, Smith 1980). The soil is characterized by its dark color, potsherds, and lithic artifacts (Hornborg 2005, Kern et al. 2004, Kämpf et al. 2003, Neves 1999, Hartt 1885) and is therefore associated with former sedentary settlements (Woods and Denevan 2009, Hornborg 2005, Woods and Glaser 2004, Smith 1980, Hilbert 1968). The diversity of ceramics indicate that at least Incised Rim, Polychrome and Incised-Punctated traditions were aware of the high fertility of Amazonian Dark Earth (Kern et al. 2004, Myers 2004, Hilbert and Hilbert 1980). The Terra Preta sites are located in various climatic, geologic, and topographic areas, usually near rivers, creeks or lakes (Woods and Denevan 2009, Kern et al. 2004, 2003). In contrast to older assumptions, current research shows, that the distribution of the sites is not limited to the bluffs along major rivers but can also be found in terra firme habitats (Oliver 2008). Since the occurrence of Terra Preta is always related to excavation sites, and therefore former settlements, an understanding of the environmental requirements of Terra Preta genesis is helpful to gain knowledge about the distribution patterns of pre-colonial cultures.

The attractiveness of the Amazon Basin as living space also influences the discussion about estimated Amazonian population density and distribution (Denevan 1996).

> We should begin by rejecting the image of Amazonia as pristine wilderness. The physical evidence alone forces us to preconceptualize the region as in some respects a cultural landscape. Studies in historical ecology suggest that more than 12% of the supposedly pristine Amazonian rain forest are anthropogenic in origin the sense that they would not exist in their present form without human intervention (Hornborg 2005, p. 590)

2.2 Spatial Analysis in Archaeology

Archaeology has always had a spatial component (Hodder 1977). The aim to retrieve information on archaeological spatial relationships as well as the examination of spatial consequences of former settlements can be seen as the key elements of spatial archaeology (Clarke 1977). Clarke (1977) distinguishes between three levels of resolution of spatial archaeology, namely the micro level (within structures), the semi-micro level (within sites), and the macro level (between sites). This research focuses on the inter-site relationships and therefore the macro level which

> comprise(s) the non-random or reiterative allocation of artefacts, resource spaces, structures and sites to particular relative loci within integrated site systems and across landscapes. (Clarke 1977, p. 13)

Whereas the author defines a site system as a set of sites where the interconnection between the sites is greater than the interconnection between sites outside the system. That encompasses the exchange of commodities and resources as well as the reciprocal movement of people.

In order to analyse the spatial factors, such as spatial distribution, intra- or intersite relationships, appropriate methods are needed. The use of visual interpretation of distribution maps has been common practice in archaeological research since the nineteen-twenties (Clark 1957), but the development of spatial studies has been slow (Baxter 2003, Hodder and Orton 1976). The development of quantitative methods for pattern analysis helps to reduce the subjectivity of the interpretation of maps (Hodder 1977, Harvey 1969, Garner 1967). That has been shown by Hodder and Orton (1976), who illustrated the possible sources of errors by comparing different interpretations of randomly dispersed points when working without statistical methods. The use of spatial statistics in order to facilitate spatial archaeology was barely used in archaeological practice until the beginning of the 21st century (Baxter 2003). Until then the work of Hodder and Orton (1976) was referenced and not superseded (Wheatley and Gillings 2002, Aldenderfer 1998).

Kantner (2008) pointed out that the differences between micro, semi-micro and macro level (Kantner named it site-focused and region-focused) are not clearly definable anymore. The author suggested to use the term region-sensitive for human-human or human-environment relationships instead. Several research fields are associated with spatial archaeology, in terms of inter-site relationships Clarke (1977) mentions settlement pattern analysis, and site catchment and territorial analysis. Settlement pattern studies gained more awareness, after Willey (1953) initiated the development of methods in American archaeology (Kohler 1988). That led to a greater understanding of settlement patterns and the environmental determinants. Inspired by that, Vita-Finzi and Higgs (1970) introduced site catchment analysis, which tries to identify the rules that determine human spatial behaviour. After the National Historic Preservation Act in 1966 cultural resource management became an important issue in America. In order to identify historic property new methods were developed in order to compare expected and observed site distributions. That was the basis for data driven predictive modelling (Verhagen 2007). Whereas predictive modelling and settlement pattern analysis have a lot in common.

> Many settlement pattern studies differ from predictive locational models only in their lack of explicit extrapolation to a spatial population. (Kohler 1988, p. 19)

The discussion about which set of variables to use for settlement pattern studies, site catchment and territorial analysis, as well as predictive modelling is still ongoing (Verhagen and Whitley 2012).

> ... we have seen that predictive modelling of archaeological resources may involve consideration of the characteristics of catchments around potential site locations, of distances to various resource types from potential locations, and of various characteristics of the potential site location itself. (Kvamme and Kohler 1988, p. 493)

Many theories exists about the predominant factors which influence the settlement location. A major distinction can be made between environmental and social factors. In case of environmental variables (Kvamme 1988) stated, that a typical archaeological approach would be to use landform or landform related phenomena and categorize them into nominal types (e.g. canyon, cliff, plain, or slope). Steepness is labelled as a relevant factor due to the assumption that steep slopes do not interfere with archaeological activities (Kvamme 1988, Roper 1979b, Judge 1973, Williams et al. 1973). Another variable is the roughness of a terrain due to the higher costs for daily activities and travel (Plog 1981, Hurlbett 1977). Rivers and lakes are seen as important not only as water source but also for transportation issues. Roper (1979a) stated that some resources, such as water, are so basic and so vital that the distance to obtain them must be minimized. Due to the importance of water a lot of archaeological settlement studies focused on the distance to different water source types (e.g. permanent rivers, lakes, springs, or seasonal streams) (e.g. Parker 1985, Scholtz 1981, Brown 1979, Lovis 1976, Judge 1973). The accessibility of resources other than water are also examined in spatial archaeology (Kvamme 1988). View is considered to be important for hunting as well as strategic purposes (Jochim 1976, Judge 1973). The protection from wind and other environmental influences is another factor for the settlement location, thus the shelter and the quality of shelter need to be factored in (Kvamme 1988, Jochim 1976). The aspect is often used in order to take shelter effect into account. Depending on the environmental conditions, the chosen exposure can protect against wind or may offer greater warmth (Kvamme 1988, Grady 1980). Zhang et al. (2014) chose elevation, slope, distance to river, distance to road, distance to coast and aspect, as several publications have pointed out the importance of these variables (Xiao et al. 2013, Zhang et al. 2013, Su et al. 2011, Goebel 2007, Gonzalez-Abraham et al. 2007). A similar selection was used by Wei et al. (2013), who factored in slope, aspect, relief degree of land surface, land use, vegetation index, hydrology and climate. Green (1973) assumed, that besides environmental properties, the distance to trade routes was relevant for the settlements and thus settlements were embedded into a larger network. Common social factors were often related to site density, site proximities, or spacing (Kvamme 1988). Whitley and Burns (2008) stated that individual decisions and the spatial knowledge should be considered. Hodder and Orton (1976) mentioned the following variables as being of importance for locating a settlement: the distance to water, the type of soil and vegetation cover, the location of other settlements, the ability to defend a settlement, the distance to suitable building materials, and the proximity of routes or roads and markets. Verhagen et al. (2013) stated that accessibility, visibility, and settlement continuity are useful in terms of predictive modelling. Other studies published similar listings (e.g. Countryman et al. 2010, Jochim 1976, Chisholm 1973), whereby the predominant factors are always related to food availability and production capacity (Zhang et al. 2014, Luck 2007, Kirch et al. 2004). The (geographical) features which define a favourable or unfavourable location can be interpreted in terms of the underlying cultural differences (Silva and Steele 2014).

The consideration of social and cultural variables for prehistoric settlement studies is very difficult and settlement strategies can hardly be described by means of coded social rules (Pizziolo 2015). Some studies tried to consider socio-cultural variables and combined environmental and social factors (Kamermans et al. 2009, Whitley 2005, Stančič and Kvamme 1999).

> But the truly determining factor for the presence of material records is the so-called śettlement pattern, a certain spatial regularity in the distribution of Human presence within a society. This regularity is itself also conditioned by natural, cultural, social and economic factors of undoubtedly deterministic nature.(Garcia 2015, p. 26)

Another aspect which is inherent in site systems is the existence of different settlement types. This means that different settlements of the same culture serve for different functions (e.g. hunting, trading or defence)

> A central assumption in archaeology is that this locations of sites of different functional categories or chronological periods will represent responses to different situational contexts, such as environmental circumstances. (Kvamme 1988, p. 329)

Kvamme (1988) mentioned temporary camps, kill sites or permanent settlement as examples for different functionalities. Brewster et al. (2003) identified three different settlement types for the prehistoric hunter-gatherer settlements in South California, namely Major Residential Bases, Limited Activity Locales, and Dinner Camps. The distance and the cost of movement to the surrounding parameters (let them be environmental or not) seem to be a major cause for the settlement location. Brewster et al. (2003) discovered, that especially the large settlement are located right next to rivers. The smaller settlements usually are a bit further away but do not exceed the distance of 700 metres to the closest river (as can be seen in figure 2.2). Binford (1982) argues that zones of economic activity exist around residential camps. A foraging radius defines the area which is exploited by the residents who return to the camp every day. A logistical radius defines the area which is exploited by a group that stays away from the residential camp but stays in (sometimes temporal) functional camps (e.g. hunting camp or fishing camp). Drury (1972) distinguishes between several types of Romano-British settlements (e.g. providing artisan and/or trading functions). A four-level hierarchy (based on the number of trenches and size of the site) of settlements in the region of Bunyoro-Kitara, Uganda, was identified by Robertshaw (1994).

Christaller (1933) presented three different principles of distribution of centres ac-

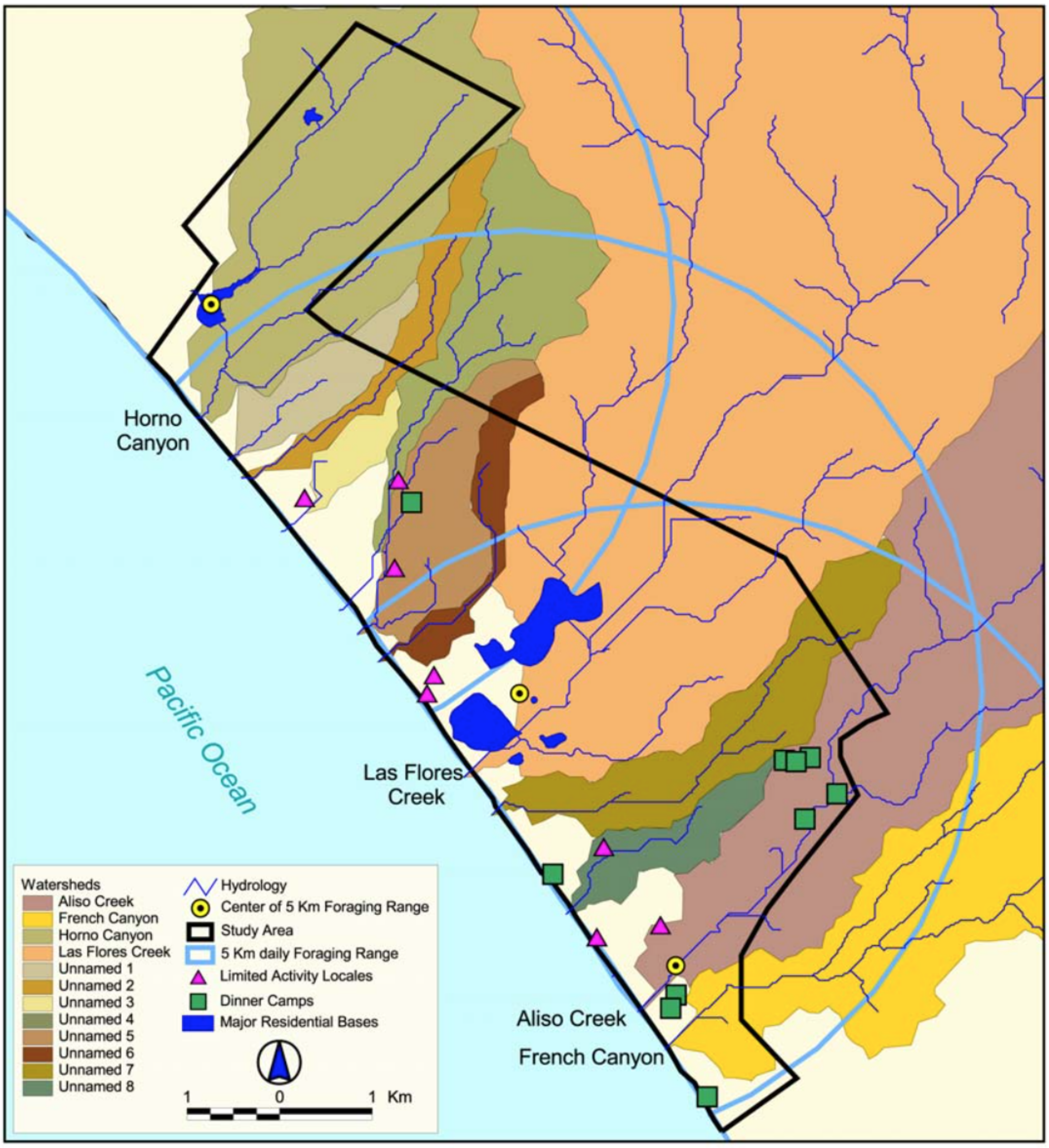


Figure 2.2: Distribution of archaeological sites by type site with drainage catchments, sites and 5-km foraging ranges for key drainages (Brewster et al. 2003)

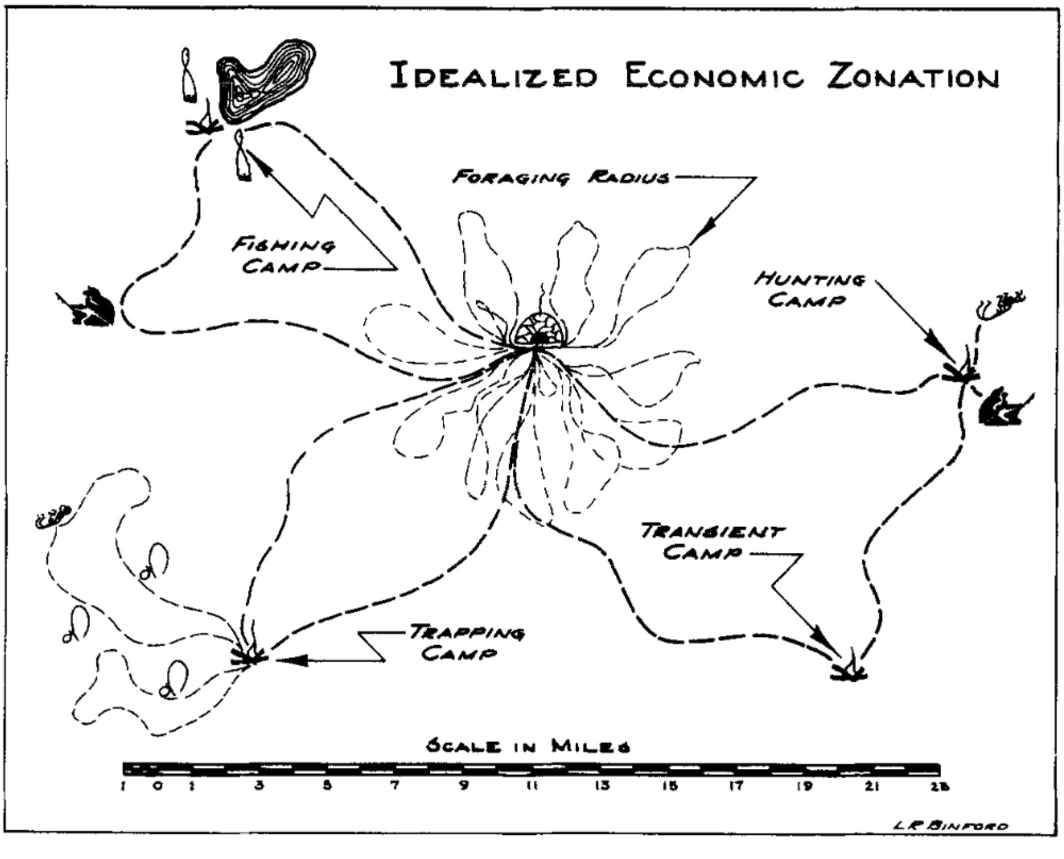


Figure 2.3: Schematic representation of zones of economic activity around a residential camp (Binford 1982)

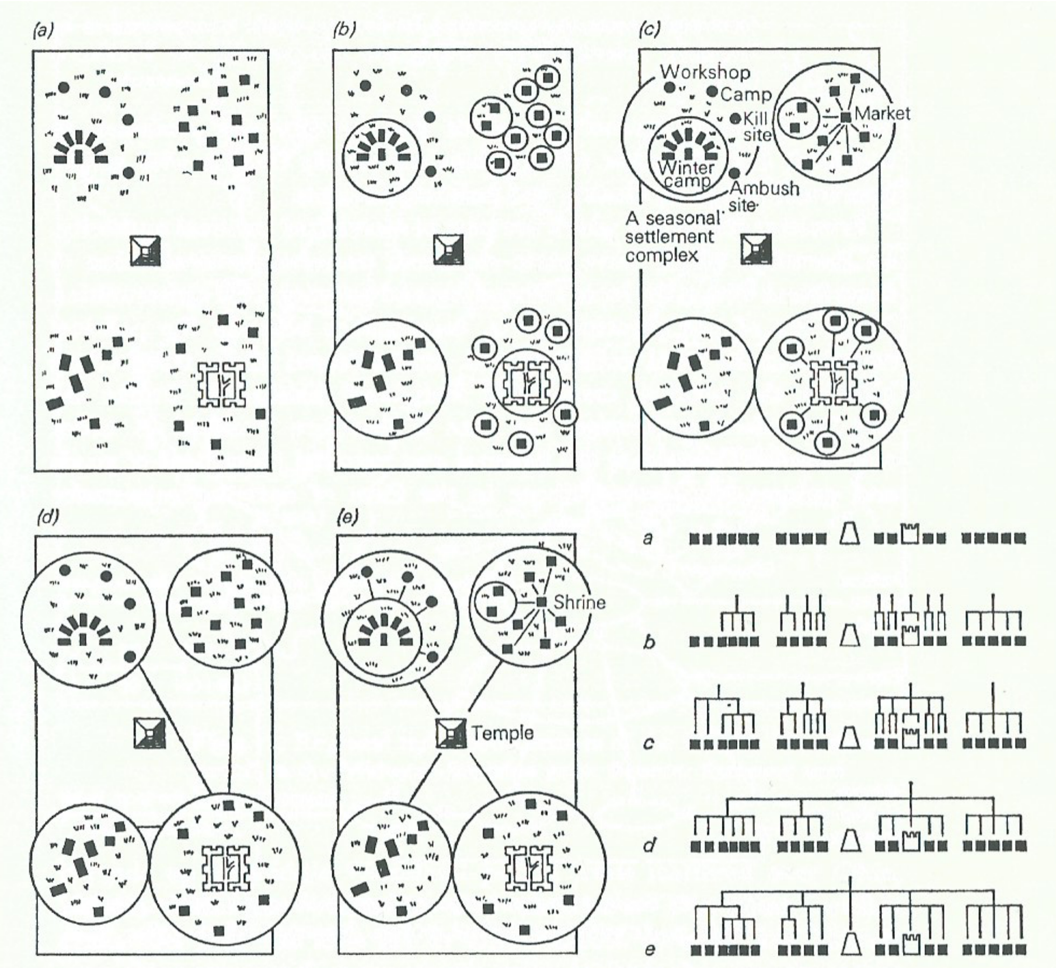


Figure 2.4: The articulation of sites into hierarchical relationships. (*a*) Archaeological ruins whose contemporaneity has been established; (*b*) spheres of domestic activity; (*c*) local networks; (*d*) political networks; (*e*) religious networks; (*bottom right*) hierarchies of the inter-relationships in *a* to *e*. (Chang 1972)

cording to the predominate principles: the market, the transport, and the administrative principle. The relationships between several sites can get complex when hierarchical relationships are factored in (as shown in figure 2.4). The service area as well as the centrality vary according to the context (Chang 1972). Regardless of the model used one aspect remains static, namely the idea that settlements do not necessarily cover the same functionalities but rather serve different needs. The following chapters describe the related spatial archaeology topics settlement pattern analysis, site catchment and territorial analysis, and predictive modelling in greater detail. A clear distinction between those topics is not always possible , especially due to the inconsistent use of those terms in the literature. A broad overview is given in the following chapters.

2.2.1 Settlement Pattern Analysis

The analysis of settlement pattern is related to spatial – and if possible – temporal properties and has always been of great interest to archaeologists.

> Regardless of their theoretical predispositions, archaeologists have always asked the simple question as to why an individual or group of individuals decided to live in a given place instead of another. (Eve and Crema 2014, p. 267)

Morgan (1881) raised the question of how the remains of former settlements reflect the social organization, which is still a core issue in settlement pattern studies (Parsons 1972). Butzer (1982) pointed out that environmental components characterize the interactions between prehistoric cultures and their environment. The location selection is influenced by geographical characteristics like topography, water accessibility or strategical purposes (e.g. Robinson et al. 2012, Su et al. 2011, Carrión-Flores and Irwin 2004, Warren 1990, Butzer 1982, Hodder and Orton 1976). In order to identify settlement patterns, two different approaches are commonly used, the social perspective and the ecological perspective. The former is used with the aim to identify behavioural groups because societies are made of interacting local groups.

> [T]ranslated into archaeology that means one has to find the places with accessible material remains of peoples' activities. (Kowalewski 2008, p. 227)

The latter approach focusses on the factors governing the distribution and abundance of a culture (Kowalewski 2008).

Willey (1953) defined a settlement pattern

> as the way in which man disposed himself over a landscape on which he lived. It refers to dwellings, to their arrangement, and to the nature and disposition of other buildings pertaining to community life. These settlements reflect the natural environment, the level of technology on which the builders operated, and various institutions of social interaction and control which the culture maintained. (Willey 1953, p. 1)

Another definition was given by Kowalewski (2008) who stated that settlement patterns are the regularities caused by the distribution of settlements or places and activities, and the relation between each other and the environment. The latter definition is used in this thesis. In settlement pattern analysis the observed settlements (in terms of spatial archaeology often referred to as sites) can either be places of permanent or temporary habitation. The analysis of the inferences from spatial distribution and environmental variables helps to understand the settlement patterns of cultures (Silva and Steele 2014, Zhang et al. 2014, Haggett 2004, Clarke 1977). Underlying is an ideological shift in archaeology towards the study of relationships between things rather than the study of things (Binford 1972).

According to Parsons (1972), the beginning of a proliferation of archaeological settlement pattern studies began in the mid-1950s. In 1955 Beardsley published the results of the so called seminars of Archaeology: 1955 – a series held by the Society of American Archaeology. This publication is what Parsons described as

> perhaps the first major effort to integrate the concept of settlement pattern within a general developmental classification of culture. (Parsons 1972, p. 129)

Settlement distributions are often described as random, regular or clustered, although these patterns rarely occur so clearly in practice (Bevan and Conolly 2006, Mayer 2006, Andel et al. 1986). Usually, human settlement behaviour is expected to be not random (Hodder 1977). That is because the settlement location is an important factor (Borsdorf and Bender 2010) which is determined by local amenities such as food availability and production capacity (Zhang et al. 2014, Borsdorf and Bender 2010, Luck 2007, Sevenant and Antrop 2007, Kirch et al. 2004, Mueser and Graves 1995). This assumption results in a distribution of settlements which can be further examined and can contribute to a greater understanding of cultural habits (Zhang et al. 2014, Willey 1953).

> It is the non-randomness which provides information about the distribution. (Hodder and Orton 1976, p. 53)

The author identified two main criteria for a random dispersion of an observed pattern:

1. a misleading aggregation of sites
2. the pattern of site survival and fieldwork

To clarify what is meant by the first aspects the distribution of hillforts in south-west England is used as an example. The pattern of the findings appear to be randomly dispersed at first sight. When the hillforts are classified according to whether they are univallate or multivallate, another pattern can be identified. This means that the classification of findings plays an important role when it comes to analysing settlement patterns. The second aspect is about the regional differences which may influence whether findings can still be found and excavated or not. Depending on soil or climate conditions some materials may endure a longer time period in one region than elsewhere. Identifying a random pattern therefore does not mean that the processes which produced that pattern were random.

According to the Central Place Theory provided by Christaller (1933), towns of similar size are uniformly distributed. Christaller (1933) stated, that some places provide goods for the daily needs and some provide specialized services, such as hospitals or universities. Assuming an isotropic featureless plain, the settlements with the same degree of centrality are equidistant from one another in order to minimize the effort (e.g. transportation costs) (Christaller 1933). Of course, this model is based on premises that do not represent the variety of potential geographical situations, yet some regularities can be identified in archaeological settlement distributions. Hodder and Orton (1976) confirmed, that in early societies markets were about 3 - 7 km apart from each other. Similar evidence was provided by Obudho (1976), Hill and Smith (1972), Smith (1971), who observed several areas in West Africa and identified a range from 5 to 15 km between markets. Drury (1972) (as cited in Hodder and Orton (1976)) showed, that certain Romano-British settlements, namely the ones providing artisan and trading functions, were all located within a certain distance range to each other. Markets too close to each other, i.e. below a critical distance, were in competition with each other and would not provide enough trade for the communities. In that case the establishment will either fail or become mobile (Fagerlund and Smith 1970). This leads to the assumption that a uniform pattern indicates some sort of competition between the sites.

A clustered pattern can occur because of various reasons. Especially in archaeology clustered patterns can also be caused due to uneven fieldwork or site survival (Hodder and Orton 1976). In contrast to the previous section, no isotropic conditions are assumed. Besides that, environmental influences such as soil type or river networks can be the reason for clustered settlement patterns. Additionally other factors such as the distance to bigger agglomerations or spiritual locations can influence the decision. Furthermore, settlements have specific needs, such as the availability of food, water, tools, building materials or other resources. However, the influencing resources vary as shown by Ellison and Harriss (1972) who studied the settlement and land use in Southern England. Examples for clustered patterns are Bronze Age barrows around Stonehenge and similar monuments or Romano-British villas around towns (Hodder and Orton 1976). According to a theory of Hudson (1969) the agglomeration may also occur because of the spread of settlement. The author distinguishes two temporal steps of spread. The first step is characterized by a small number of settlements in order to colonise an area. This pattern may even be randomly dispersed. In a second stage of spread, additional settlements occur due to the increasing population. These settlements are located close to the initial center with the tendency to only move short distances to keep distances short. That process leads to a clustered settlement pattern. To examine this model of spread, Wood (1971) suggested the use of probability distributions. To estimate a probable distribution, the observed area is subdivided into squares. Based on the total number of observed settlements, the probability of settlements being within one square is calculated. The comparison of the observed pattern with the expected pattern allows conclusions about the applied model. A couple of so called contagious models were used whereas the term contagious indicates an increasing probability of an occurrence if nearby occurrences exist. Davey (1971) analysed the distribution of bronzes and identified clustered patterns with regard to natural regions using a χ^2 test in order to examine the goodness of fit. Kruk and Machnik (1973) (as cited in Hodder and Orton 1976) also used a probability function to analyse the settlement patterns of the early neolithic in the south Polish loess zone. The authors noticed, that in two of three analysed periods (the authors used the term periods but it seems to be similar to the term culture which is used in this thesis), the clustered pattern of settlements is the result of the functional inter-relationships of sites, which confirms the theory of contagious spread. In the third period settlements appear to be randomly dispersed with the tendency to appear closer to rivers. Kruk and Machnik (1973) assumed, that this pattern was the result of ongoing spread. The problem of using quadrat counts in order to estimate the probability distribution is, that the size of the squares has an impact on the result. Harvey (1968) stated that the size of the square needs to cover a cluster but should not include more than one which also means that the clusters need to be at a sufficient distance from one another.

Another common approach is the use of nearest neighbour functions in order to identify a settlement pattern (i.e. Linard et al. 2012, Tian et al. 2012). Nearest neighbour functions measure the distance(s) to *n* nearest neighbours and thus refer to the position of the surrounding settlements (Clark and Evans 1954). One problem of this method is the potentially misleading influence of edge effects. Another problem can be the lack of independence, which means that the correlation is 1.0 when two points A and B are each others nearest neighbours (Dixon 2012, Cox 1981) - so called isolated or mutual nearest neighbours (Schilling 1986, Pickard 1982). These problems can be addressed by using edge correction and factoring in *n* nearest neighbours.

Bevan and Conolly (2006) criticised that the common approaches focus on neigh-

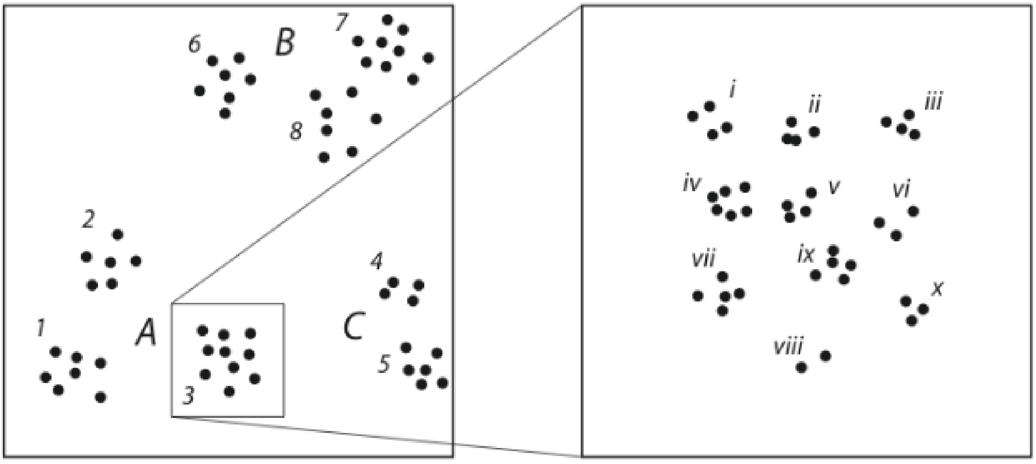


Figure 2.5: Multiscalar point patterns (Struever and Houart 1972)

bour distance functions and thus are unable to detect patterns on different scales. Therefore, the authors present a multiscalar approach in order to detect patterns on different scales. Figure 2.5 shows an example of 56 hypothetically distributed settlements. Applying a K-means statistic would suggest that the optimum number of clusters would be eight. Taking a closer look at the left panel of figure 2.5 shows that there are another 3 clusters (A - C) on a higher-order scale. By including the resolution of the artefacts at an excavation site (see right panel at figure 2.5) another scale and thus a third pattern is shown. The authors use the Ripley's K-function which allows the analysis of point processes at different scales, further allowing a multi-scalar analysis. That is an advantage over the commonly used nearest neighbour function which only detects patterns on one scale (Diggle 2014, 2003, Bailey and Gatrell 1995).

The Ripley's K function is still commonly used for settlement pattern analysis. Palmisano (2013) applied the function in order to distinguish between what the author called first order and second order of effects of settlement patterns. The properties of the local environment are defined as the first order of effects whereas the processes based on interaction between settlements are defined as the second order of effects. Zhang et al. (2014) wanted to analyse the spatial characteristics of a settlement location rather than only the area, density or shape of settlements. The case study took place in eastern coastal China. The authors also used the Ripley's K function in order to analyse the point patterns of human settlement. Besides the Ripley's K function a Monte Carlo simulation was included in order to get more reliable results with a quantifiable statistical significance. To identify geographical determinants of the settlement locations the authors employed a regression model. A stepwise multiple linear regression was performed where the settlement pattern category, namely random, clustered, or regular, was used as the dependent variable. The authors conclude that the settlement pattern relies on the location of the settlement, meaning that clustered patterns more likely tend to occur in interior, hilly areas and regular and random patterns can be found close to the coastal plain. Fan et al. (2009) used remote sensing data to compare spatial and temporal expansion patterns. A similar approach is used by Liu et al. (2005) who used Landsat images to identify changes in land use. Li et al. (2011) tried to derive an environmental suitability for human settlement by selecting and analysing natural factors such as the terrain, hydrology, or vegetation using remote sensing and GIS. The use of remote sensing data in order to compare the changes in land use are only applicable for the last decades. In terms of archaeological settlement pattern analysis remote sensing can be used in order to identify new settlements – e.g. due to the detection of different soil conditions, vegetation types, or topographical characteristics – but not for some sort of pattern finding. Kirch et al. (2004) tried to identify geographical influences using linear regression. The authors used the elevation as the dependent variable and thus ignored any spatial relationships. (See chapter 3.2.1 for a detailed explanation of the functions mentioned above).

Another option is the use of agent based models which allows the consideration of human experiences and changing environmental conditions (Paliou 2008). Heckbert (2013) published a model of the ancient Maya social-ecological system. This includes the change of spatial landscape due to climate variation or anthropogenic impact. Another (earlier) example was given by Axtell et al. (2002) who tried to model the population growth and collapse of the Kayenata Anasazi culture. Both models were able to more or less reproduce the varying spatial pattern over time.

In a review article about regional settlement pattern studies Kowalewski (2008) compared settlement pattern analysis according to their behavioural pattern. The author classifies the pattern based on the predominant social structure into categories such as chiefdom, villages and communities, or states. Each category goes along with specific arrangements of inter-site relationships and specific arrangements of site locations. Thus, it can be assumed that conclusions about human settlement behaviour can be drawn by the analysis of site location pattern.

2.2.2 Site Catchment and Territorial Analysis

Other research fields on macro level are site catchment, and territorial analysis. At some sites (especially in humid conditions) plant and animal remains are poorly preserved. This means that any assumptions about the economies and agricultural origins are complicated (Renfrew and Bahn 2005). In order to be able to draw conclusions, the environment was factored in. The term site catchment analysis was initially introduced by Vita-Finzi and Higgs (1970) in order to describe the analysis of archaeological sites and the relation to their environmental surroundings (Renfrew and Bahn 2005). That means that not only artifacts but also the possibilities inherent in the location of the site and its surroundings are relevant (Vita-Finzi and Higgs 1970). The authors defined it as

> the study of the relationships between technology and those natural resources lying within economic range of individual sites. (Vita-Finzi and Higgs 1970, p. 5)

Thus, a catchment of an archaeological site is defined as the area which is needed in order to derive its resources (Roper 1979b). The aim is to emphasize parameters such as the availability or abundance of resources such as plants, animals, or materials as determinants of the location (Roper 1979b). Roper (1979b) used the term man-land relationship in contrast to, e.g. the Central Place theory which focusses more on distances or population density and is therefore labelled as man-man relationship. However, the distinction between those two relationships is not unique and borders are blurred. It is assumed that the distance to the necessary resources plays an important role for the site location. There is a close relationship between the amount of energy that is needed for procurement of resources and the distance that must be overcome.

> Human populations are generally only able to exploit resources that exist within a certain distance of their occupation site, be this a camp, cave, village or town. (Jarman 1972, p. 706)

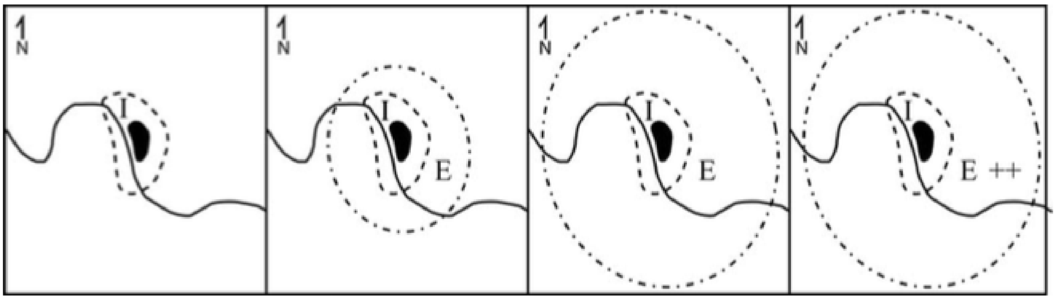
Vita-Finzi and Higgs (1970) stated, that an area is more likely to be exploited the closer it is to the site. Henshall (1968) presented several research papers that support the assumption, that distance is directly related to the settlement location (for Africa by Prothero 1957, and Fortes et al. 1947; for India by Ahmad 1952; for Brazil by Waibel 1958). Lee (1969) observed the !Kung in southern Africa and found out that they usually do not go further than 10 kilometres to procure resources. Chisholm (1973) assumes, that the critical distance normally is even less than 10 kilometres.

> It is further assumed that prehistoric peoples were aware of this decrease in cost/ benefit ratio and located sites, moved their locations, and generally played out a settlement strategy that minimized the ratio of energy expended to energy procured. (Roper 1979b, p. 121)

Another assumption is that a hierarchy of resources exists. That means that people are willing to cover greater distances for some resources than for others (Roper

1979b, Jochim 1976). This lead to a division into different zones of the area around a site whereas the zones cover different resources and thus activities. Von Thünen (1875) published a model of concentric rings of land use which would occur around a city. It focusses on agricultural activity and the distances of specific land use patterns. Although the model is based on several premises, such as the existence of an isolated state, and is therefore only a very simplified description of the reality, it reflects some important aspects. The model was later modified in order to consider other factors influencing the land use. These factors covered other competing markets, differences in the productivity as well as rivers, in order to facilitate the transportation. Even though this model was created in the late 19th century and therefore is not tailored to archaeological problems, it was found to be applicable for archaeological needs (Roper 1979b, Hodder and Orton 1976, Chisholm 1973). It is often used in site catchment and territorial analysis because of the consideration of distances and the usage of zones based on land use (Renfrew and Bahn 2005). The size, shape and location of a specific site provides information about the settlement strategy (Christopherson et al. 1999). Steward's (1938) study about aboriginal groups showed that different environmental conditions correlate with different settlement strategies. That leads to a basic premise of site catchment analysis: The location and the site function correlate and inferences about the function can be made when the location is known (Roper 1979b).

An important distinction is made between catchment and territory. While a territory is the area which can immediately be accessed, a catchment is the total area from which contents of a site are obtained (Higgs 1975, Vita-Finzi and Higgs 1970). Based on this Higgs and Vita-Finzi (1972) suggested to distinguish between site catchment analysis and site territorial analysis, whereas these techniques are complementary. Site territorial analysis is a more theoretical approach using maximum walking distance or time values (e.g. 3 hour walking distance or 5 kilometres). In contrast, site catchment analysis is about locating the nearest source of the materials which were actually found in the excavation site and therefore a more empirical approach (Renfrew and Bahn 2005). These distances delineate a maximum movement radius. Conclusions about the mode of subsistence can be drawn by analysing the available natural resources within the radius (Ducke and Kroefges 2007). Sites can have more than one catchment which can differ from each other and also from the territory (e.g. due to trading relations the economic catchment area can be larger than the exploited territory). These two approaches are often mixed up and a lot of the published research on site catchment analysis is in fact site territorial analysis (Renfrew and Bahn 2005, Roper 1979b).



Population needs could be met by intensive garden strategy within 500 m of settlement	Population needs could not be met by intensive garden cultivation alone, but possible with limited extensive agriculture	Population needs could be met only by mixed intensive cultivation and extensive agriculture utilizing fields up to the 4 km radius	Large populations at centers required surplus from other villages

Figure 2.6: Different sizes of gardening catchments (Duffy 2010)

In order to identify a sites' territory, fixed radii or time contours are used (Tiffany and Abbott 1982, Styles 1981). Other methods are used in order to derive site catchments. Findlow and DeAtley (1974) identified two types of sites and measured the distance between the settlements. Afterwards these distances were combined with the distance along and across drainages. The observed spacings were used to derive size and shape of catchments. A similar approach was used by Browman (1976) who tried to determine catchment sizes in Peru. Flannery (1976) tried to determine the catchment area by looking at the findings (plants, animals, and mineral resources) and tried to identify their origin. Duffy (2010) simply created a buffer around the sites in order to derive the site catchment area. The author used different distances to consider different population sizes (see figure 2.6). Cassels (1972) argues that

> the most likely boundary between the two sites is a line equidistant between them. (Cassels 1972, p. 215)

Several other studies adopted this approach because of the geometric simplicity (e.g. (Angell and Moore 1984, Danks 1977, Hodder and Orton 1976)). Some problems occur when a territory is derived by the use of Thiessen polygons. First of all, no geographical information is considered. Which means that no natural barriers or hydrographic features are taken into account. The results may be misleading when contemporaneity of sites cannot be assumed. Another problem is the potential inaccuracy due to an incomplete set of excavation sites. Additionally the results can be biased due to the extent of the study area. However, it is a commonly used technique for the segmentation of space into territories (Renfrew and Bahn 2012, Ducke and Kroefges 2007, Conolly and Lake 2006, Wheatley and Gillings 2002, Cormack 1979). Conolly and Lake (2006) tried to overcome some of the shortcomings and additionally considered site weights to get better results. Ducke and Kroefges (2007) criticised that Thiessen polygons still have drawbacks in terms of site territorial analysis. Thiessen polygons are always complete which means that there is no area which is not assigned to a territory. There is no measure of error because it always provides perfect partitioning. The partitioning usually results in a division into several pieces with crisp boundaries (which is not always the case). The shape and size of the polygons is very sensitive to changes in the set of sites. An alternative to the Thiessen polygons is provided by Renfrew and Level (1979) who developed the XTENT model. The model is based on the size of the site as well as the distance to other sites. This is in contrast to the Central Place Theory and other approaches where the level of hierarchy determines the size of the territory (Renfrew and Bahn 2012). In the XTENT model it is assumed that the size (or importance) of each site is directly proportional to its area of influence. This leads to bell or bell-tent shaped areas where the height of the tent is related to the importance of the site and the influence decreases with an increasing distance to the settlement. If an associated bell is completely covered by a larger area of another site it is considered to be subordinate (Renfrew and Bahn 2012, Ducke and Kroefges 2007, Renfrew and Level 1979). The XTENT model was defined with the following equation:

$$I = f(C) - k * d \tag{2.1}$$

The strength of an area of influence I is determined by two variables, namely the center size (the weight of a site) C and the distance d. By subtracting these two values a large center near to another site has more influence (means having a high I value) than one far away, but a large center can still be dominant. Renfrew and Level (1979) worked with

$$f(C) = C^a \tag{2.2}$$

but theoretically other functions are also possible. The coefficients a and k allow the user to adjust the two variables. The importance of size increases exponentially, the importance of distance increases on a linear basis (Ducke and Kroefges 2007). Other assumptions underlying the XTENT model (besides the correlation of size and influence) are the following. The territories are continuous and not spatially interrupted, a piece of land is uniquely assigned to one site or territory and capitals of an area of influence have a higher weight (usually population size) than subordinate sites (Ducke and Kroefges 2007, Renfrew and Level 1979). There is only a small number of studies which applied the XTENT model (e.g. Soetens et al. 2003, Hare 2001, Grant 1986, Scarry and Payne 1986), because the model is very sensitive to changes in the coefficients a and k and needs to be calibrated subjectively for each study (Ducke and Kroefges 2007). Ducke and Kroefges (2007) modified the XTENT model in order to consider hierarchical relationships and maximum territorial reach in advance which makes it easier to control the results. Geographical features such as the topography are not incorporated into the mathematical model (Grant 1986, Renfrew and Level 1979). In order to derive more realistic results, the distance measure could also be replaced by a cost of movement measure (Conolly and Lake 2006, Wheatley and Gillings 2002).

> Sets of distances have often been calculated [...] as a first approximation, as Great Circle distances (geodesic, in the original sense of that word). However, if we wish to explore the possibility that geography influenced directions and rates of spread from some origin point, then we must obtain sets of distances that reflect the influence of geographical features. Silva and Steele (2014, p. 611)

Cost surface techniques are most commonly used to obtain such distance estimates (e.g. Field et al. 2007, Glass et al. 1999). Ullah (2011) calculated catchment area(s) on the basis of anisotropic travel costs derived from a cost surface. Lönnqvist et al. (2009) calculated the cost surface based on satellite imagery using Tobler's hiking distance (Tobler 1993). This function allows to determine the hiking speed taking the slope into account. Once a site catchment area is estimated it can be analysed. Vita-Finzi and Higgs (1970) defined land use capability classes such as irrigated land, arable, sand dunes, or rough grazing. The percentage of occupation by each land use class can be calculated within time contours or distance buffers around each site and thus allows conclusions about the subsistence pattern (see figure 2.7 for an example) (Dibyopama 2010, Pappu and Shinde 1990, Pappu 1988). Other approaches used multivariate statistics, such as cluster analysis or factor analysis, to facilitate the interpretation (Roper 1979a, 1974 as cited in Roper 1979b).

2.2.3 Predictive Modelling

The term predictive modelling in archaeological research can be traced back to the 1970s but became more popular at the end of the 1970s and the beginning of the 1980s (Verhagen 2007, Baxter 2003, Sebastian and Judge 1988). One aim of predictive modelling is the identification of variables which distinguish between sites and non-sites (Baxter 2003). Another aim is to predict the probability of archaeological remains at a specific location (Verhagen 2007). It allows researchers to

> formulate expectations about the future state of a system that are based on our knowledge of such systems or similar ones (Sebastian and Judge 1988, p. 2).

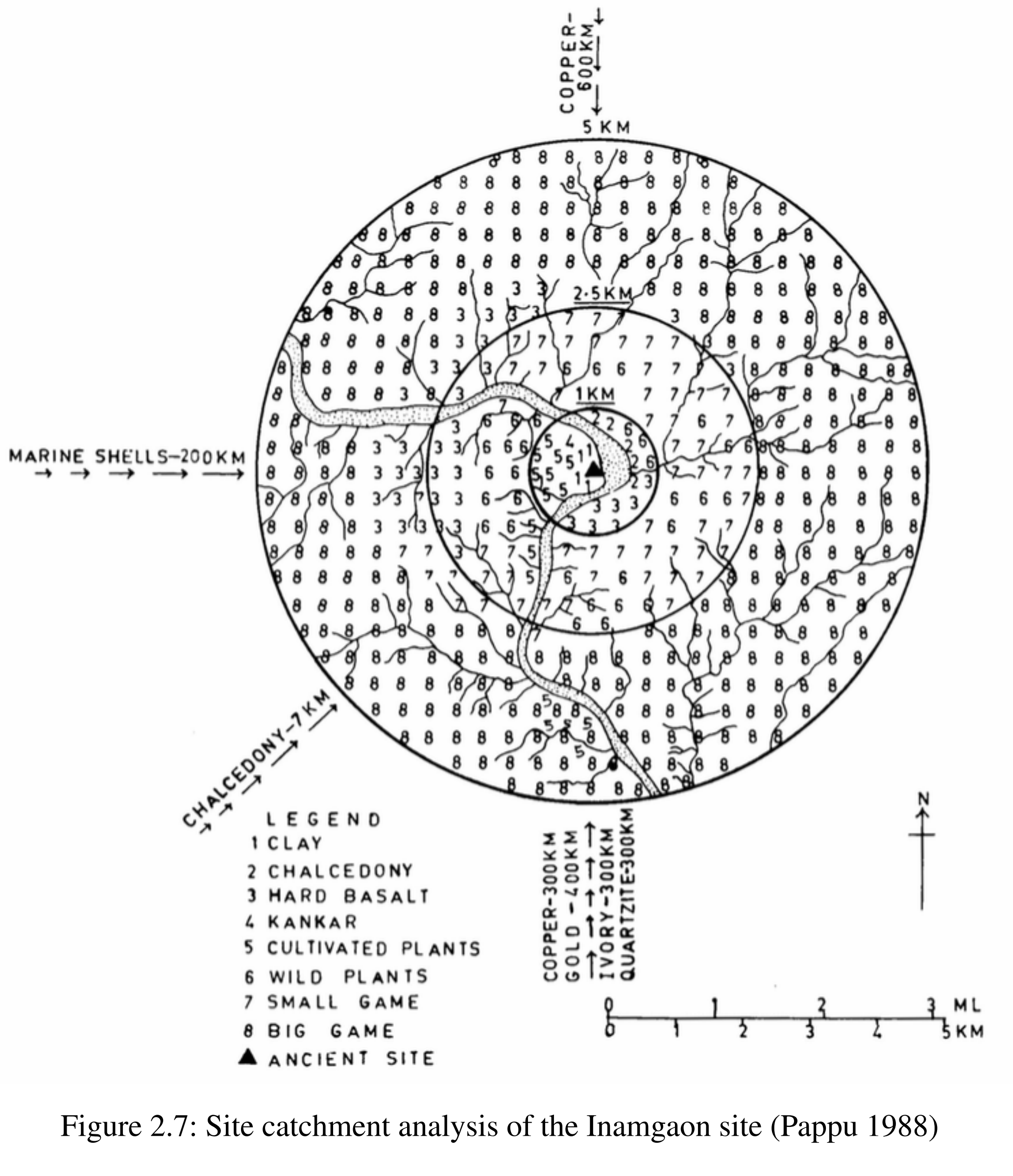


Figure 2.7: Site catchment analysis of the Inamgaon site (Pappu 1988)

Predictive modelling is based on the assumption that human settlements are not randomly dispersed but rather related to certain characteristics (Verhagen 2007). A definition was given by Kohler and Parker (1986) who stated that predictive modelling tries to predict

> at a minimum, 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 and Parker 1986, p. 400)

Whereas the term predictive might not be clearly defined in general, it clearly indicates the anticipation of the spatial distribution of archaeological records in terms of spatial archaeology (Garcia 2015). Casarotto (2015) proposed to use the term location preference modelling rather than predictive modelling. The author argues that due to inherently vague data, the aim should be to test assumptions, hypothesis and possible scenarios of ancient settlement and not so much a prediction of those. Sebastian and Judge (1988) used the terms correlative and explanatory models in order to distinguish between models based on a sample of the region and models based on human behaviour. The first is a more inductive approach which identifies and quantifies relationships between settlements (also see figure 2.8) and the surrounding environment whereas the latter is more deductive and derives models based on the understanding of human behaviour and cultural systems (also see figure 2.9). Correlative models can be used to predict where sites might be but are insufficient for the understanding of subsistence strategies (Kincaid 1988, Sebastian and Judge 1988). The reason is, that correlations can be detected but no information about causality is gained. Explanatory models in return are used in order to understand cultural variability and similarity or the cultural stability and change. The major limitation of explanatory models is the difficulty in creating and validating such models. In many cases a deductive predictive model is not more than an idea which can be tested against the observed data (Kamermans and Wansleeben 1999). Hempel (1970) stated that the aim of predictive modelling should be to derive statements describing certain conditions as well as suitable general laws. These general laws could then be used in order to predict other locations of prehistoric activity. Gilman and Thornes (1985) assumed that minimal cost models are the cause of locational patterns. If that was true a direct relation between settlement location and environment would exist. The authors applied an analysis of variance in order to verify this hypothesis. Land use potential (irrigated, dry farming or unfarmed were used as categories), the distance to other settlements (within or without a 30 minutes walking distance), rainfall levels, topography and the chronological-cultural period were considered. This is one of the examples which provided satisfactory results (Garcia 2015). Most of the time inductive approaches are used because it is im-

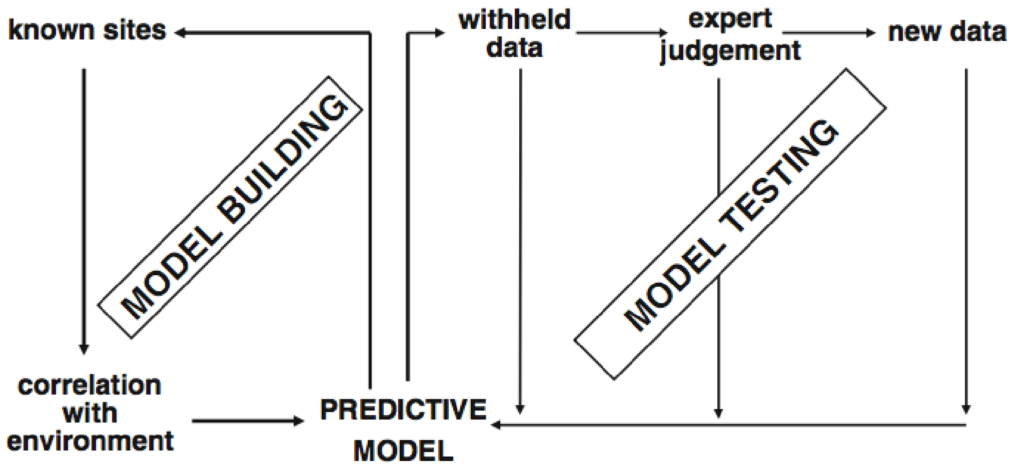


Figure 2.8: The procedure used for inductive predictive modelling. A statistical comparison of archaeological data and "environmental" variables is used to create a predictive model. This model is then tested either through statistical methods using withheld or new data or by means of peer review (expert judgement). (Verhagen and Whitley 2012)

possible to formulate a deterministic hypothesis (Garcia 2015, Berrocal 2005).

Initially, research focused on the development of statistical and spatial analysis methods in order to perform data driven predictive modelling (Verhagen 2007). At a very simple level the location or land parcel is the unit of analysis, which is defined as either being an archaeological site or not (Kvamme 1988). More statistically spoken, a dependent variable (in this case of type boolean) is assigned to the unit of investigation. Either simple environment categories or – which is usually more sufficient – complex multivariate functions can be assigned in order to predict the probability of settlements (as shown in figure 2.10). Models can be considered as suitable if the number of correct predictions is higher than with a random model (Kvamme 1988). Kvamme (1988) used the following equation in order to measure the quality of a model and to compare different models.

$$Gain = 1 - \left(\frac{\text{percentage of total area covered by model}}{\text{percentage of total sites within model area}}\right) \tag{2.3}$$

Based on this equation a predictive model performs best when a quite small area is likely to contain sites and most of the found sites are within that area. This leads to the following: with a gain statistic close to 1 a model has a good predictive use whereas a value close to 0 indicates that the model has little or no predictive use. If the value is below 0 the model has reverse predictive use which can still be of some use. In this case the areas outside the specified areas are considered to be the

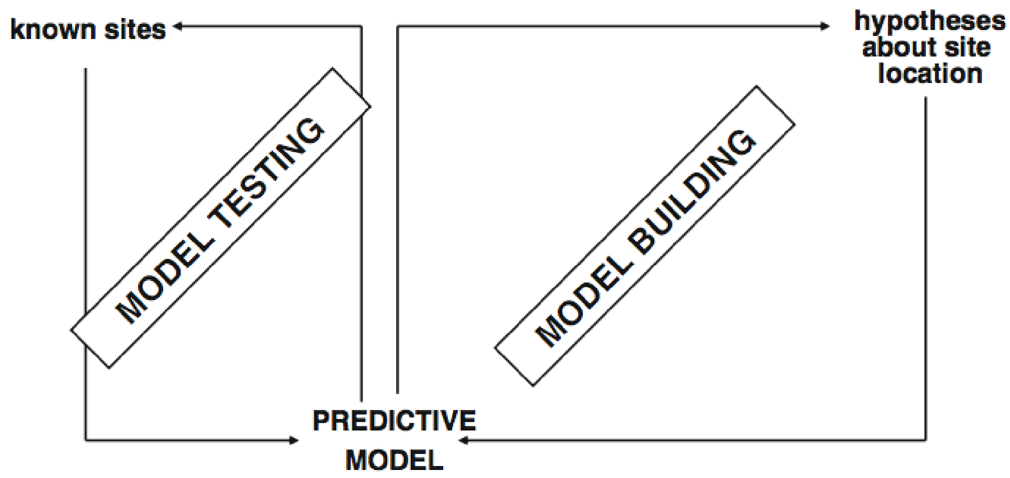


Figure 2.9: The procedure used for deductive predictive modelling. The model is created on the basis of hypotheses on site location preferences. In many cases, these are no more than "educated guesses". The model can easily be tested with statistical methods as the archaeological data set is not used for building the model. (Verhagen and Whitley 2012)

areas being modelled. In contrast to simply measuring the percentage of correct prediction – which can be misleading – this gain statistic facilitates the comparison of models (Kvamme 1988).

It can be assumed that archaeological sites fulfil different functionalities – such as residential, hunting, or trading – (e.g. Kvamme 1988, Binford 1982, Hodder and Orton 1976) which may require different situational contexts. Establishing different site-type models can be challenging in practice due to difficulties in the assignment of the type. Often, site-types are assigned based on limited and sometimes questionable evidence (Kvamme 1988). To avoid incorrect site-type assignments Kvamme (1985, 1983) suggested to measure the quantity of archaeological activities and investigate why certain locations were used in the past and others not.

A lot of different approaches exist in order to perform predictive modelling with logistic regression being the most widespread (Wheatley 2004, Warren 1990). One commonly used technique is the trend surface analysis, which uses the observed locations in order to derive trends (Hodder and Orton 1976, Unwin 1975). Roper 1976, Monroe et al. 1980 and Bove 1981 applied this procedure to model a continuous surface of site distribution across a region. Trend surface analysis is a regression function which tries to estimate the properties at any location based on the properties of the known locations (Unwin 1975). The analysis is usually about the presence or absence of sites or site-types or other nominal scaled variables. This can be

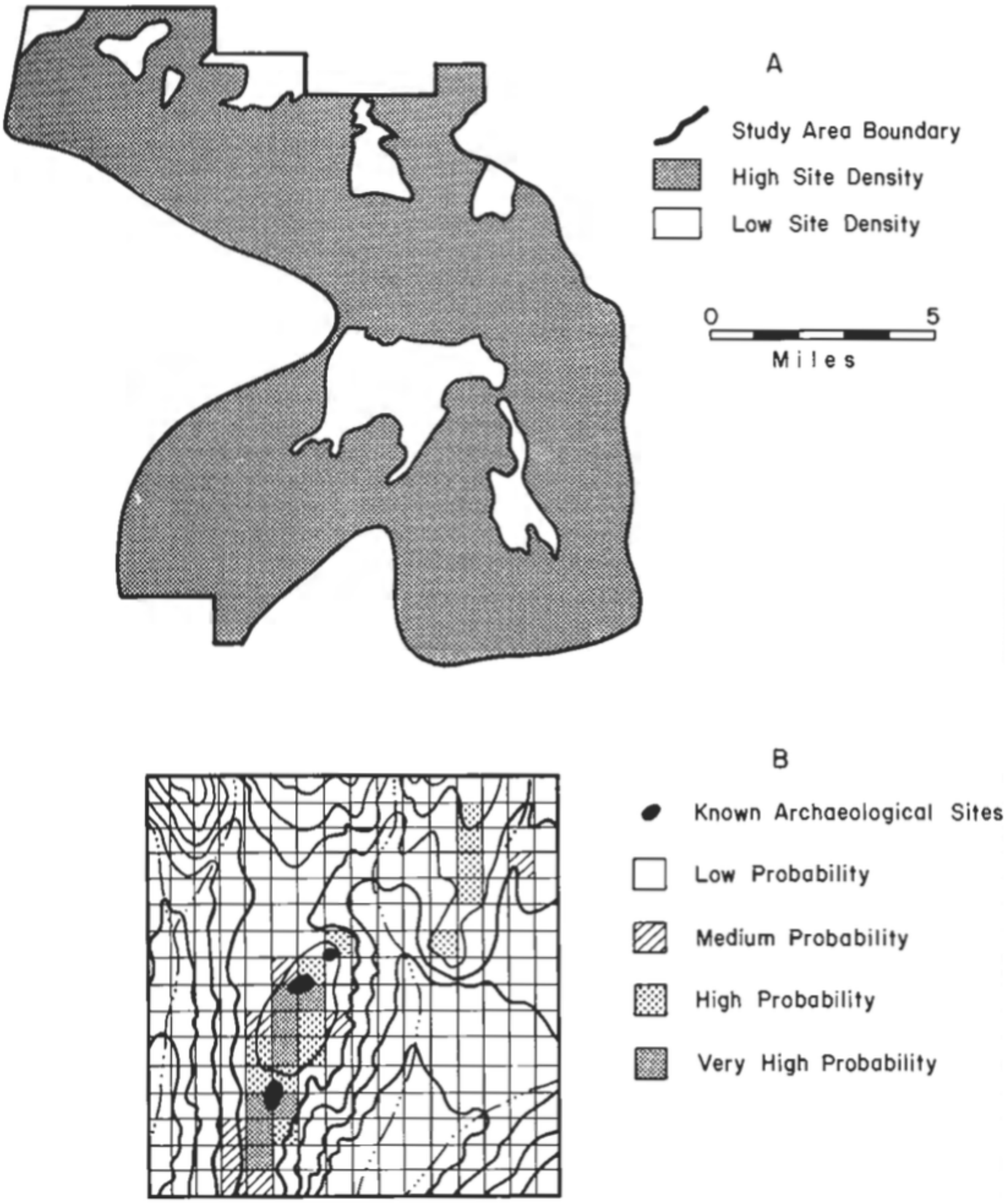


Figure 2.10: End products of cultural resource modelling (A) A simple plant community mapping in which the communities correspond to different site densities (after Plog 1983, p. 64). (B) A "site probability surface" superimposed on a map and derived from a complex multivariate function of six variables measured in each 50 by 50 m cell (after Kvamme 1980). (Kvamme 1988)

problematic as most regression analyses need numerical scaled dependent variables (Kvamme 1988). Another problem is the applied grid size because it influences the resulting trend surface. Kriging, which is another interpolation technique considering spatial variance seems to provide better results (Kohler 1988, Parker 1985, Zubrow and Harbaugh 1978). In order to allow nominal-level categories Wrigley (1977) developed a logistic trend surface analysis which makes no assumption about the distributional form (e.g. normal distribution etc.). Depending on the complexity of the distribution either a linear regression function (first order) or a polynomial function (2nd, 3rd,...,*n*th order) can be used. Detailed information about the procedure can be found in the boxes 1 – 3 in chapter G.

In order to consider locational characteristics, which are often on a nominal level, Grady (1980) subdivided the observed area into smaller parts and projected the expected number of sites based on the density estimates derived from observations. A similar approach was realized by Judge (1973) who focused on continuous site location information, which allows to consider distances to specific environmental variables (e.g. x percent of all sites are within a distance y to a river). This approach allows the consideration of spatial information. Especially since GIS became more popular, the number of predictive models has been increasing (Casarotto 2015). A commonly used approach is the reclassification of spatial information such as soil type or slope in order to derive a land units map (Casarotto 2015). The reclassification is recommended for some of the environmental parameters (e.g. it is usually more important if the soil is fertile or not than the exact name of the specific soil). An important aspect of the analysis is to test the data against sources of bias. Leonardi (1992) suggests to create a map of geochronological units in order to identify areas were findings are possible or not. If landslide, erosion or other environmental influences took place archaeological evidence might have disappeared or is less likely to be found. By considering such aspects a predictive map can be created showing the suitability of locations for former settlements. Rogers and Curdy (2015) used cost surfaces and least cost paths in order to narrow down the areas of potential prehistoric activity in the Pennine Alps. By determining the possible mountain pass routes the archaeological observation areas can be minimized. Egeland et al. (2010) also used a cost surface to calculate the least cost path to determine probable other settlement locations. Countryman et al. (2010) derived a suitability surface using weighted overlays of seven variables which the authors hypothesized to be influential. A subsequent field survey was used to revise the former model. Stančič and Kvamme (1999) stated that for small sample sizes boolean overlays should be used because robust multivariate statistical methods cannot be applied. One major disadvantage of this technique is its inability to quantify the influence of each variable. Whitley and Burns (2008) criticised that usually only environmental influences are considered when a probability surface is calculated. Hence, the authors defined activity-based settlement/subsistence patterns which shall represent common prehistoric occupations. Based on these categories specific weights can be assigned to input variables and thus allow a detailed analysis which fits the preferences of the observed sites. Other approaches are based on artificial neural networks (ANN). This concept is inspired by biological neural networks, in particular the brain. It belongs to the artificial intelligence techniques because of its ability to learn (so called adaptive models) (Abdi 2003, Abdi et al. 1999). ANNs are built from (several) units, often referred to as neutrons due to the analogy to the brain. Those units are linked by weighted connections, whereas the learning process is often accomplished by modification of the weights (e.g. Deravignone and Macchi Jánica 2006a, Abdi 2003, Wang 2003, Abdi et al. 1999). A commonly used architecture in ANN is the multi layer perceptron (see figure 2.11) which is similar to multivariate non-linear regression techniques in a statistical framework. A certain number of units are interconnected from the input to hidden units and from the hidden units to the output. The training pattern teaches the ANN how to get a specific output with a certain input (Deravignone and Macchi Jánica 2006a). This architecture is able to handle fuzzy data because of its ability to use logic paradigms. It allows to define thresholds and thus regulates the flow based on rules involving one or more units. Deravignone and Macchi Jánica (2006b) applied this concept to develop the Grosseto Predictive Modeling Method, which initially focussed on the districts of Grosseto, Siena and Arezzo but was also applied to other regions (Blankholm 2015). The idea is to use the observed presence/absence pattern as output and environmental, social and/or economic variables as input. The authors developed a tool which works as a bridge between GIS and ANNs to consider spatial influences.

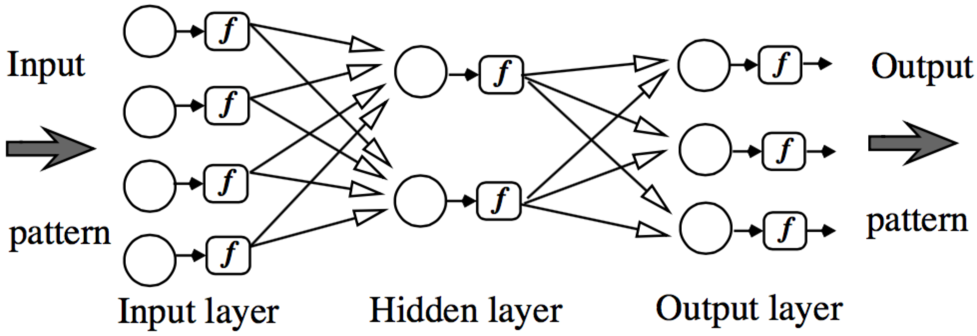


Figure 2.11: A multi-layer perceptron (Abdi 2003)

Another statistical approach was applied by McMichael et al. (2014), who used the maximum entropy model (MEM) in order to predict the occurrence of terra preta in the Amazon Basin. The aim of maximum entropy modelling is to approximate an unknown probability whereas that approximation should be used which satisfies best any constraints that the analyst is aware of (Phillips et al. 2006, Jaynes 1957). MEMs seem to outperform other statistical approaches especially in the case of presence only datasets (e.g. Elith et al. 2011, Phillips and Dudík 2008, Elith et al. 2006, Phillips et al. 2006). Presence only datasets only consist of (more or less) reliable observed data whereas in contrast presence-absence data also consist of data which indicate absence. That absence information can be misleading because it is hard to obtain reliable data (Jiménez-Valverde et al. 2008) and thus may preclude modelling of potential distributions (Svenning and Skov 2004). While there are a couple of studies in terms of biology and the behaviour of species there are only a limited number of archaeological applications. Rivers et al. (2011) tried to determine networks of Bronze Age Aegean civilizations in order to describe flows between sites, goods, ideas etc. The authors used the maximum entropy model to identify the *most likely* networks. McMichael et al. (2014) tried to identify potential terra preta findings in the Amazon Basin by applying the maximum entropy model. See chapter 3.2.2 for a detailed explanation of MEM.

Predictive modelling is still subject of controversial debate (e.g. Van Leusen et al. 2005, Wheatley and Gillings 2002, Wansleeben and Verhart 1997). Reasons are the inherently vague and incomplete archaeological records and the (sometimes highly) subjective use of environmental factors. This is also related to the data quality because usually no historical environmental data is available. Another reason is the problem of environmental determinism which is often inherent in predictive models (Wheatley 2004, Kvamme 1997). In contrast Verhagen (2007) stated that indeed the predictive models are simplistic but can fulfil their need and raised the question if a predictive model necessarily needs to have full explanatory power. Several studies aim to reduce the potential area rather than increase model accuracy (e.g Blankholm 2015, Rogers and Curdy 2015).

3 Methodology

The analysis of functional settlement pattern requires several steps. The main input for the analysis is the location as well as the surrounding environment. Consequently the coordinates of the former settlements – if not yet known – and the environmental variables need to be determined. Whereas the former is mainly done based on literature research, the latter requires further analysis. Besides the location of the former settlement, the storage of this data is a relevant task in order to facilitate research in terms of Amazonian archaeology. A database on a server and a data model which meets the requirements of the archaeological dataset is created. It can be assumed that not all environmental variables need necessarily to be analysed (e.g. air pressure or carbon dioxide content). Additionally, some of the data is missing for the observed period. Another aspect is the classification of environmental data (e.g. soil types can be used individually or can be aggregated based on soil properties such as fertility). Therefore the available environmental data is analysed based on the relevance for the observed excavation sites. A first overview can be gained by plotting the measured values for all locations and identify anomalies. Afterwards the Maximum Entropy Modelling (see chapter 3.2.2 for a detailed explanation) allows to determine the variable importance for the selected settlements. Another preparatory step is the analysis of the spatial distribution of the observed locations. The identification of functional settlement pattern requires a clustered pattern. A test for complete spatial randomness (see chapter 3.2.1) is used to check whether the settlements are randomly, clustered or regularly dispersed. If this precondition is true functional settlement patterns can be calculated based on the locations as well as the environmental variables. This is done using a consensus clustering approach (see chapter 3.2.3) which compares the results of several cluster runs in order to determine an optimal cluster solution. Those results can then be used for the calculation of the suitability surface. Therefore the consensus values for each settlement function in combination with the determined measures for the environmental variables are combined and used to define a fuzzy membership function (see chapter 3.2.4). The suitability surface is calculated for each settlement function individually and form the basis for the final calculation of the territory. The maximum cumulative suitability value within a predefined distance is determined and used as limiting value for the territory (see also chapter 3.2.4). An overview of the methodology is given in figure 3.1.

3.1 Archaeological Record

In order to draw conclusions about the functional settlement pattern the surrounding environment was analysed. Settlement covers a spatial as well as temporal aspect due to seasonal fluctuations, expulsion or other reasons. Unfortunately, temporal information (usually radiocarbon measurements) is only available for a limited number of excavation sites. Thus, the approach focusses on spatial analysis in order to determine environmental variables. Several input datasets, which are listed and explained in the following chapter, were used to achieve that goal. The input can be split into two different types. On the one hand, the archaeological record, namely the locations of the excavation sites and the settling culture(s), on the other hand the surrounding environmental parameters which may determine the settlement location. The environment is considered because it is known that different cultures developed different survival or subsistence strategies (e.g in the late Holocene some cultures developed slash-and-burn as well as semi-intensive agricultural strategies). Depending on the culture various domesticated and semi domesticated plants were known Heckenberger and Neves (2009). Some cultures were able to use complex techniques such as wetland management or fish farming (Schaan 2004, Erickson 2000).

> Diverse agricultural strategies were coupled with systems of faunal exploitation that included a variety of managed species, such as birds (Muscovy ducks, parrots and macaws, and others), fish and other aquatic species, including the giant Amazon river turtle (up to 80 cm) and manatee, or sea cow. (Heckenberger and Neves 2009, p. 253)

Thus it can be assumed that the influencing environmental parameters vary according to the capabilities and needs of the culture.

The localization of former settlement locations was a time consuming task due to the absence of usable digital sources. The process of geocoding of known excavation sites is described in greater detail in chapter 3.1.1. A detailed description of the used database schema which was developed in order to store the findings is given in chapter 3.1.2.

3.1.1 Identifying and Locating Excavation Sites

Archaeological data is inherently vague which in this case also implies that not all the former settlements are known. The archaeological record can be biased due to

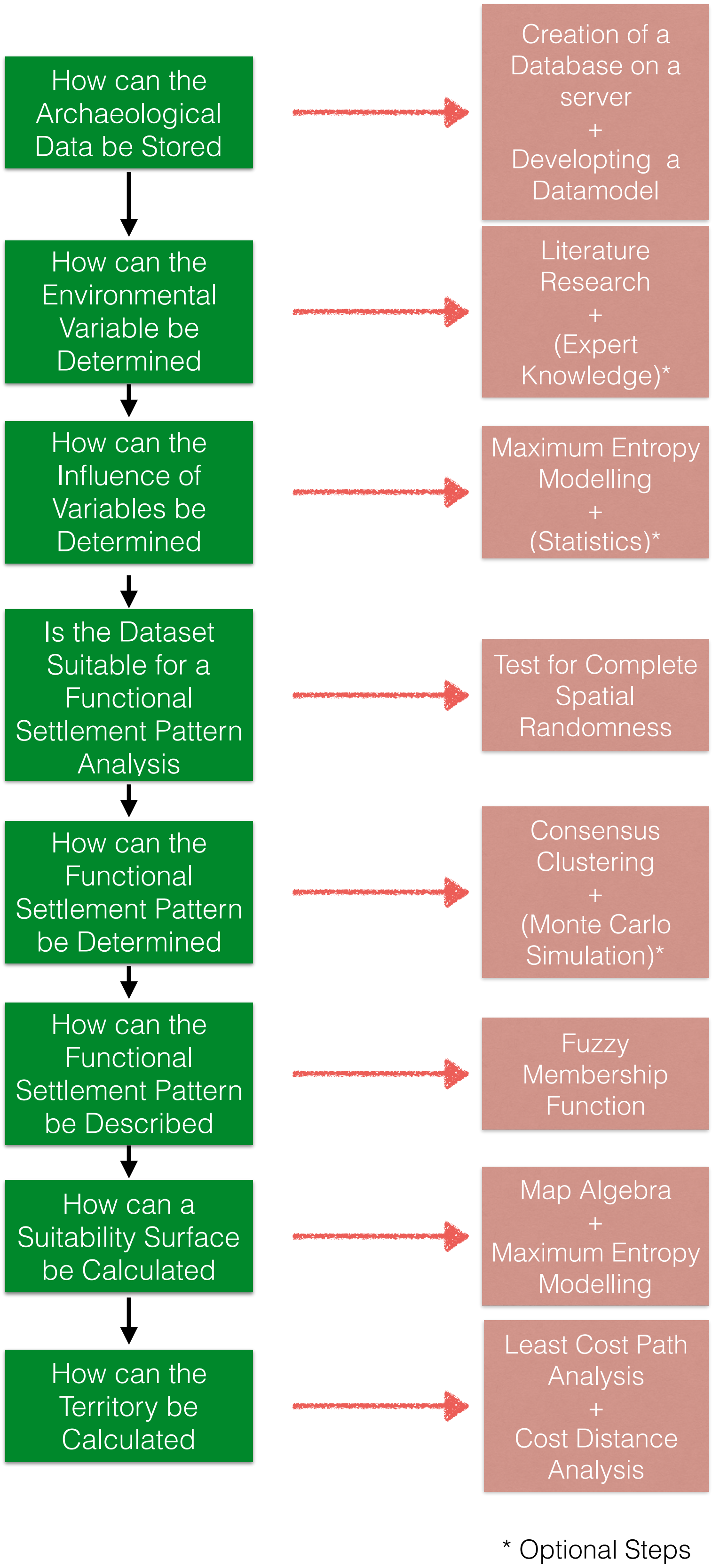


Figure 3.1: Overview over the key questions and the applied method(s)

lack of knowledge (e.g. the settlement was in a floodzone and therefore no findings are available), absent data sources about the former settlement (caused by the high number of publications in Portuguese, German, Spanish, French, etc. and due to missing central databases or institutions to administrate or collect these information) or missing location information (usually the coordinates are missing because the exploration of the excavation sites was done before GPS was commonly used). Lots of archaeological studies in the Amazon focus on a certain (small) area (or a specific excavation site) in the Amazon and analyse them in greater detail (e.g. Nunes et al. 2013, Bozarth et al. 2009, Nunes 2009, Py-Daniel 2009, Rebellato 2007, Machado 2005 for the Hatahara site close to Manaus). These studies on micro-level are used to extract the location and cultural information but cannot be used in terms of understanding a settlement patterns on a smaller scale. In order to geocode as many excavation sites as possible, over 60 publications of several research fields (e.g. archaeology, soil science) were analysed in order to get the location information. Some of those contain GPS coordinates whereby most of these publications only provide maps and descriptions of the specific sites. Additionally, old maps and reports of the chroniclers helped to relocate some of the excavation sites which are destroyed or no longer accessible (e.g. due to dam constructions and reservoirs). However, some of these locations are only an approximate. Based on the description (or illustration on a map) of the location in combination with expert knowledge (in this case Klaus Hilbert, who is an archaeologist and specialist for Amazonian archaeology) the former location of the settlement is estimated. This was done using google earth, which is easy to use and provides recent as well as historical satellite imagery which facilitates the geocoding process. Due to the availability of older satellite images, location errors due to construction projects (e.g. dams) and flooding can be minimized. Some problems occurred due to the used scale. Some authors only used small scale maps which only approximate the location of a (former) settlement (as shown in figure 3.2). Others provided maps with a higher resolution, which help to identify the position (as shown in figure 3.3. Mário Ferreira Simões (Simões and Lopes 1987, Simões and Kalkmann 1987, Simões and de Araujo-Costa 1987, Simões and Machado 1987, Simões and Corrêa 1987, Simões 1983, Simões and Araújo-Costa 1978) and Corrêa, Conceição Gentil (Corrêa 1987) published almost one third of all located excavation sites in the Amazon. The authors provided a map and a detailed description (an exemplary description of the site AM-DE-01 is given in quotation below of the location which helped to facilitate the geocoding process. The quotation can roughly be translated as follows: T̈he former settlement of the ceramic phase Cuaru is located at the left riverside of the rio Negro below the mouth of the river Demeni. It comprises an

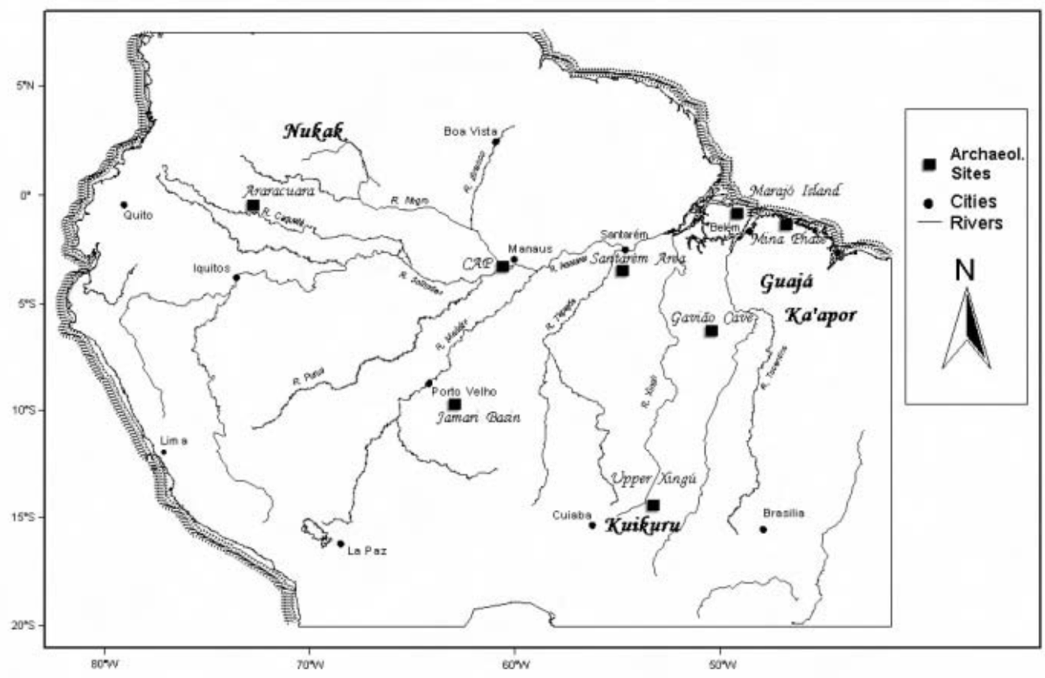


Figure 3.2: Locations of sites in the Amazon (Neves et al. 2003)

area of 250 * 100 metres parallel to the river bank and is 2 metres above the river (in september). It is important to point out that the location can be flooded in the rainy season. The publications of Mário Ferreira Simões are an important source because the author aimed for a very detailed, consistent and structured description of the sites.

> Sítio-habitação de fase cerâmica Cuaru, localizado à margem esquerda do rio Negro, abaixo da foz do rio Demeni. Compreende uma área de 250 * 100m, paralela à margem do rio, com 2m de altura em relação ao nível do rio (setembro), o que importa ser alagada nas enchentes excepcionais do rio. (Simões 1983, p. 23)

Additionally some geographical conditions are considered. Variations in vegetation and land use can indicate former settlements due to the fertile soil Terra Preta (an anthrosol). Older clearings are a sign for fertile soils which have been used for farming for generations. Figure 3.4 shows the outline of the former settlement at the Hatahara site. Based on the documentation of chroniclers in the Amazon it is further assumed that the settlements were close to water, not deep in the forest and usually which huge range of vision but not always visible (e.g. Denevan 1996). Another indicator for a former settlement is an existing settlement. Based on this information, a potential settlement location is derived which is used for further analysis. This time consuming work was necessary due to the absence of any useful digital data

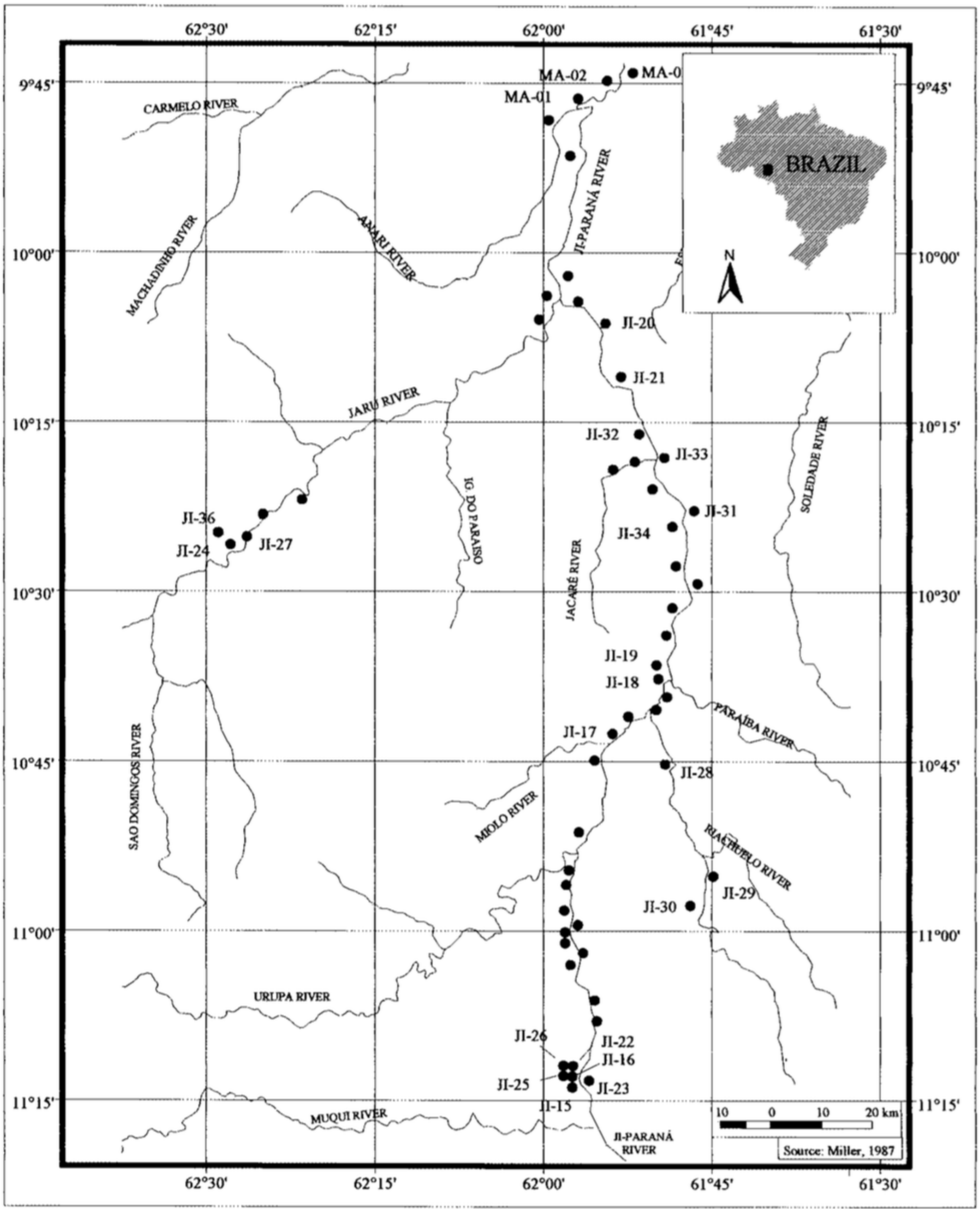


Figure 3.3: Map of the distribution of archaeological sites in the region of the Ji-Paraná river (Rondonia State) (Kern et al. 2003) based on (Miller and Caldarelli 1987)

Figure 3.4: The red line outlines the habitation zone of the Hatahara site (Rebellato et al. 2009)

sources. There is a database which is hosted by the IPHAN (link can be found in chapter 1) which provides a huge collection of excavation sites for Brazil. Besides the information about the cultures which settled there a unique excavation site name is set which is also used in archaeological publications (and thus needed in order to match the information with the location). Unfortunately no location information (coordinates) is given which makes the database useless for spatial analysis. Pardi (2002) published a map which shows archaeological sites according to their ceramic tradition. This map cannot be used to relocate the excavation sites manually because it only provides information about the tradition but not their cultural subgroups (for a detailed definition of the terms tradition and culture see chapter 2.1.1), the resolution is not good enough and the density of excavation sites in the Amazon Basin is very low.(WinklerPrins and Aldrich 2010) tried to establish a, what the authors call, *interactive GIS of Amazonian Dark Earths* which is basically a google kmz file which can be downloaded and that provides information about the location (e.g. the expected data quality, the size, age (if available), etc.) and about the data source of the points (as shown in figure 3.5). The included sites are not explored by the authors but are the result of a literature research articles (mainly the ones which are used in this geocoding process as well) and contains (state 18.08.2014) 505 Terra

Preta sites. Unfortunately the published link seems to be outdated because the website is no longer available. The authors tried to distinguish between high, medium and low quality, based on the expected location accuracy level of the excavation site. Although the locations are enriched with additional data, neither information about the finding (and thus settled cultures and traditions) nor the unique excavation site name is given. Therefore the two data sources provided by IPHAN and (WinklerPrins and Aldrich 2010) cannot be matched. Another problem that was noticed is that the data quality - even of the "high"-labelled features - seems to be not very reliable. Simões 1983, 1978 often referred to adjacent settlements which facilitate the verification of a location. Whereas other settlements in the kmz file are located several kilometres apart of the estimated location. This assumption is underlined by the fact that some settlements in the dataset from (WinklerPrins and Aldrich 2010) are several kilometres away from any water source which – based on the archaeological literature and the reports of chroniclers – does not seem to be very plausible. However, this is yet the only (other) known attempt to provide an as complete as possible collection of Amazonian Dark Earth sites. Other approaches such as the use of remote sensing techniques (Menze and Ur 2012) or Maximum Entropy Modelling (McMichael et al. 2014) in order to identify potential settlement locations are not considered. These approaches aim to predict settlement locations but it is not known whether a settlement existed at that location or not. For now the database contains 665 excavation sites with 750 findings whereas a finding is not defined by single artifacts, (which might at first be intuitive) but rather by distinct cultures. If the database stores three findings for one excavation site it simply means that this settlement was recolonized by another culture (for a more detailed explanation of the database schema see chapter 3.1.2). The data set contains 81 different cultures and 14 traditions and they cover a time span of almost 6000 years. One can assume, that capabilities and needs vary according to the observed culture/tradition and also to the time of occurrence.

After analysing the various data sources it turned out, that all excavation sites can be attributed to 19 archaeologists who are referenced by the other authors. Over 96% of all located sites were excavated by only 11 archaeologists, almost one third of all excavations – 183 sites – were supervised by Mário Ferreira Simões (see figure 3.6 and for detailed 3.7 information).

3.1.2 Storage of Excavation Sites

The located excavation sites and additional data about the excavation sites are needed in order to analyse settlement patterns. To facilitate further analysis this data needs

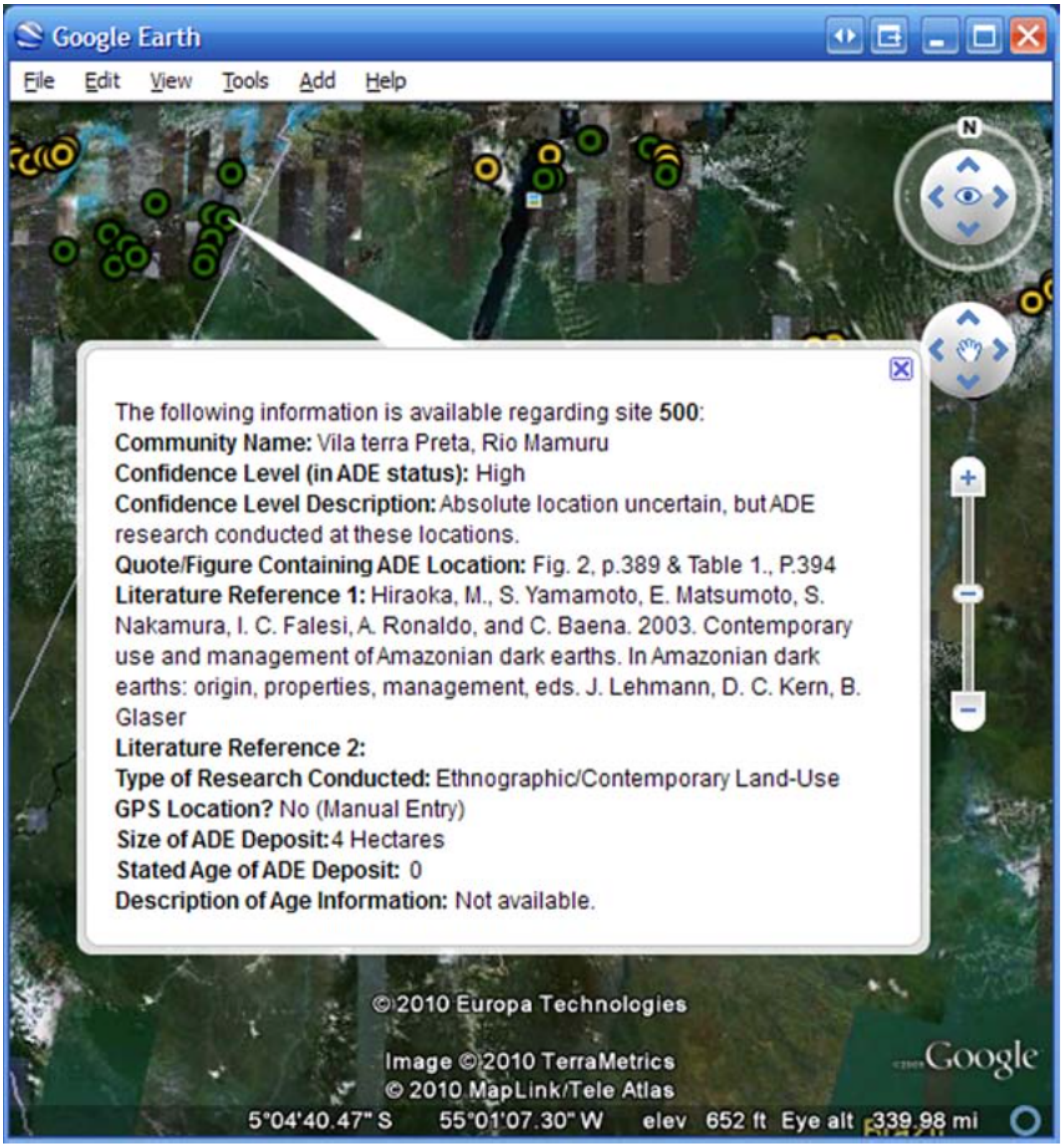


Figure 3.5: Example of the provided information for an excavation site in the kmz-file (WinklerPrins and Aldrich 2010)

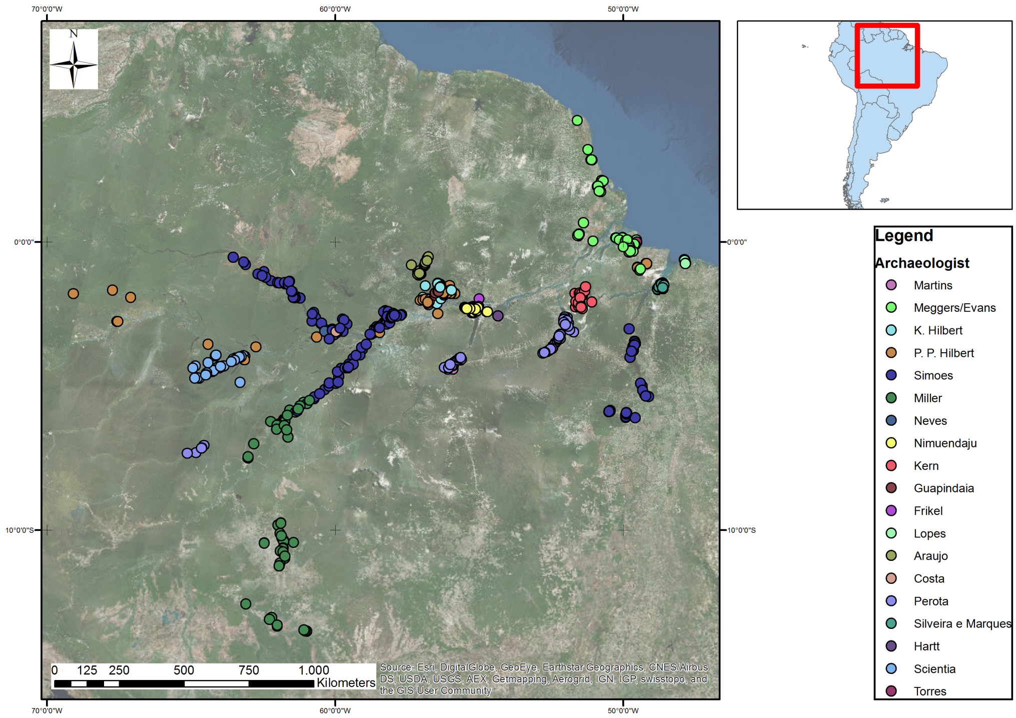


Figure 3.6: All located excavation sites color coded by the supervising archaeologist

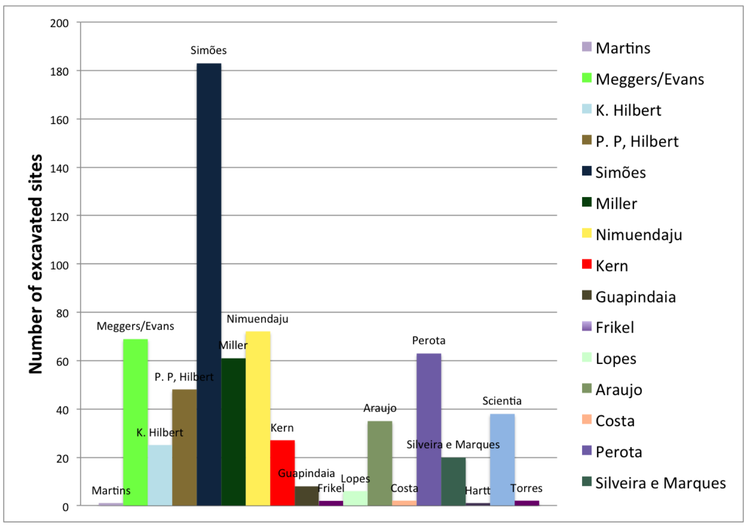


Figure 3.7: Frequency of supervised excavations by archaeologist

to be converted into a spatial data format (such as shapefile, geodatabase, other spatially enabled database formats, etc.). In this case a spatially enabled database is used, which provides data on a server and thus makes the use of local storage devices redundant.

To keep it simple, a table (e.g. Excel file) is used at first to collect the coordinates and information about the excavation sites, thus no GIS or database knowledge is needed. An excerpt of the structure of the used Excel file is shown in table 3.1. The original Excel file contains the following columns:

- Excavation Site ID
- Name
- Archaeologist
- X-coordinate
- Y-coordinate
- Length
- Width
- Depth
- Tradition
- Culture
- Radiocarbon date
- Literature source

In order to provide a simplified overview, some of the fields are not shown in table 3.1. Information for the fields length, width, depth and radiocarbon date are only available for a little number of excavation sites, which is the reason why they are not shown in the table. Nonetheless they are imported in the database. This (Excel) file was imported into the database using a python routine. This script creates the database entries based on the predefined database schema. The database schema is tailored to the needs of storing excavation sites and its formerly resident cultures. It is subdivided into nine tables for it to be normalized and to minimize data redundancy. Some of the information is directly taken from the input Excel file whereas other information is calculated and written in the database afterwards.

An excavation site can be populated multiple times (so called multi-phased vs.

Excavation Site ID	Name	Archaeologist	y	x	Tradition	Culture
PA-M-7	Aberta	Meggers/Evans 1948-49	0° 7'44.24"S	49°43'54.49"W	Incised-Punctated	Aruã
PA-OR-88	Acabaxi	K. Hilbert	1°32'30.48"S	56°17'54.59"W	Incised-Punctated	Konduri
AM-AN-01	Acará	P. P. Hilbert 1959	3°38'14.39"S	62°45'24.26"W	Polychrome	Guarita
AM-BO-06	Acará	Simões 1981	4°22'23.17"S	59°39'36.73"W	Polychrome	Guarita
AM-MC-19	Acará	Miller 1979	6°13'31.31"S	62°14'32.93"W		Marmelos
PA-M-3	Acauan	Meggers/Evans 1948-49	0° 4'10.46"S	49°37'25.30"W	Tupiguarani	Acauan
AM-IR-02	Açutuba	G. Neves	3°5'42.018"S	60°21'56.1384"W	Polychrome	Guarita

Table 3.1: An overview about the located excavation sites and the available information

single-phased if only one settling culture is known). Due to that and in order to avoid redundancy, the excavation sites and the findings are stored in separate tables. The table *excavation_sites* contains optional site relevant information such as the unique excavation site id, depth, length or width and some foreign keys which reference other tables. The table *findings* contains only foreign keys; one which indicates the excavation site itself and additionally two foreign keys which store the culture and tradition. The distinction of cultures respectively traditions in the Excel file is made based and ceramic styles. However, the database schema is designed so that any kind of distinguishing features can be used. Thus, the approach is independent from the archaeological discourse about the categorization of findings. This allows the usage of language groups or other similar traits as classifier. The tables *cultures* and *traditions* are designed in the same way – both provide a column for the name as well as for the known dates of existence (from and to). The relationship between cultures and traditions is hierarchical which means that one culture is assigned to exactly one tradition whereas one tradition usually is assigned to several cultures. A link table is used to store this hierarchical relationship. Another table is used to store the radiocarbon data. The radiocarbon data is usually provided in a format such as $year \pm year$ (e.g. 1000 ± 50), which is the reason why the plus/minus value is stored in the database. Additionally, the foreign key of the finding (because the radicarbon data is assigned to a culture, not to an excavation site) as well as a foreign key to the laboratory which analysed the sample is stored. Three tables are used to store the metadata such as the radiocarbon laboratory, the archaeologist who excavated the site as well as the literature source. An entity relationship model of the database schema is shown in figure 3.8. A detailed overview of the data model is given in figure 3.9. As mentioned above, using a database which runs on a server allows to access the data from the internet, which is used to facilitate further settlement patterns analysis. A web-based GIS for archaeology in the Amazon is embedded which provides a small toolbox for further exploratory analysis – a beta version is already available at `http://terrapreta.geo.uni-augsburg.de`. The data format as well as the spatial reference is set in advance. The excavation sites can be selected (either manually or based on the culture or tradition) for further analysis and the distance to some environmental variables can be compared in order to identify pattern or differences. The visualizations are an important communication tool because the data can be provided so that archaeologists (and other groups of interest) can use them without necessarily being an GIS expert.

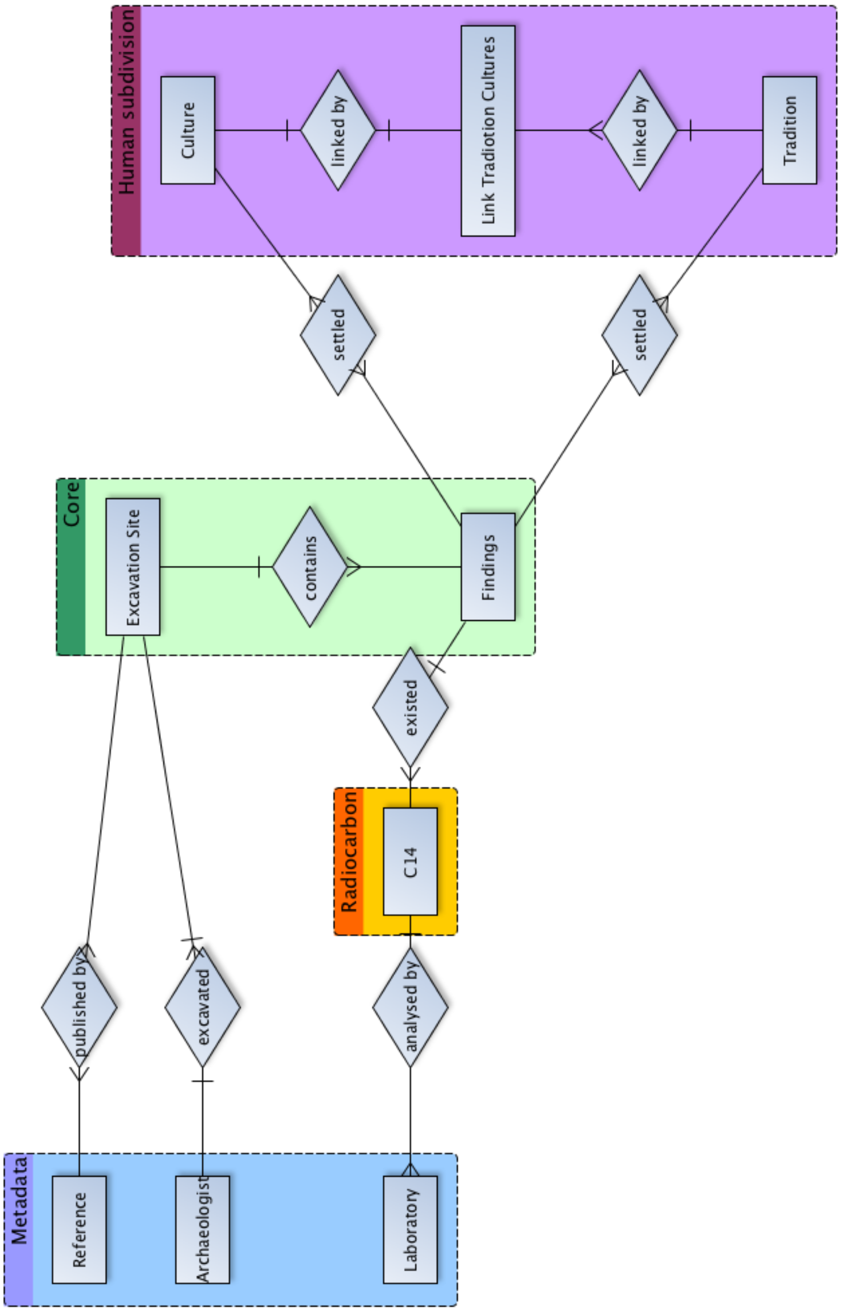


Figure 3.8: Entity relationship model of the archaeological record

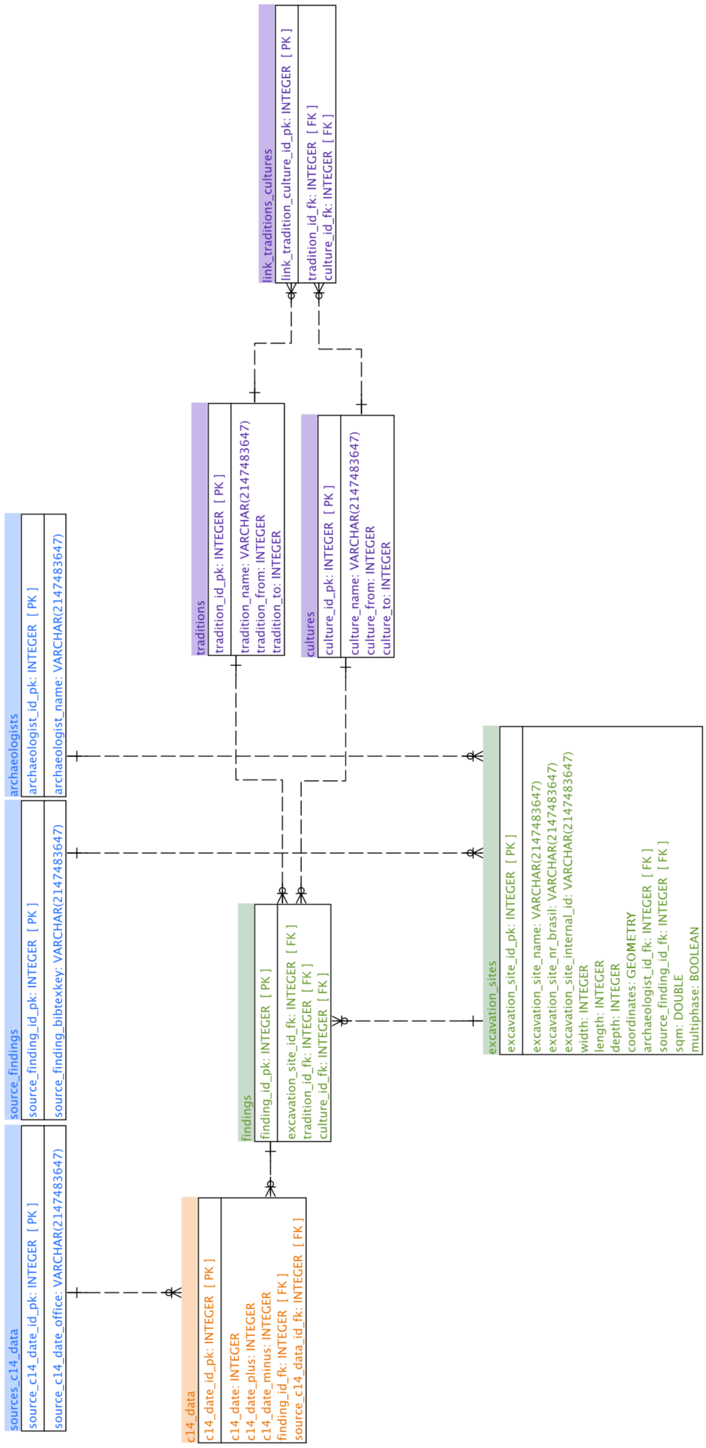


Figure 3.9: Detailed data model which is designed to fit the requirements of the archaeological data

3.2 Functional Settlement Pattern Analysis

Due to technological and computational improvements various digital datasets are now accessible (Guo and Mennis 2009, Miller and Han 2003, Norton 1999, Fayyad et al. 1996b). However, the increasing number of digital information limits the applicability of traditional analytical methods which are designed for small and homogeneous datasets (Miller and Han 2003). This leads to the development of new methods and algorithms to manage and analyse such datasets and their interdependencies (Guo and Mennis 2009, Norton 1999).

> There is often much more information in these databases than the 'shallow' information being extracted by traditional analytical and query techniques. (Miller and Han 2003, p. 4)

Thus, the extraction of useful information and knowledge is an important task in order to analyse such complex relationships (Fayyad et al. 1996b). An automatic, exploratory approach for modelling and analysing these large datasets is called knowledge discovery from databases (KDD) (Maimon and Rokach 2006b). In contrast to traditional statistical analysis KDD is more strongly inductive (Miller and Han 2003).

> Statistical models are confirmatory, requiring the analyst to specify a model a priori based on some theory, test the hypotheses and perhaps revise the theory depending on the results. (Miller and Han 2003, p. 5)

Although knowledge discovery processes usually aim to validate a hypothesis, they can also be used in order to discover new patterns autonomously (Fayyad et al. 1996b and see figure 3.11).

KDD is interdisciplinary by nature and benefits from intersecting research fields, e.g. machine learning, pattern recognition, databases, or data visualization (Norton 1999). Whereas the

> unifying goal is extracting high-level knowledge from low-level data in the context of large data sets (Fayyad et al. 1996a, p. 39)

The terms KDD and data mining are often used synonymously but data mining is in fact only one component in a larger process (as shown in figure 3.10). A commonly excepted definition of KDD is provided by Fayyad et al.:

> KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. (Fayyad et al. 1996b, p. 40)

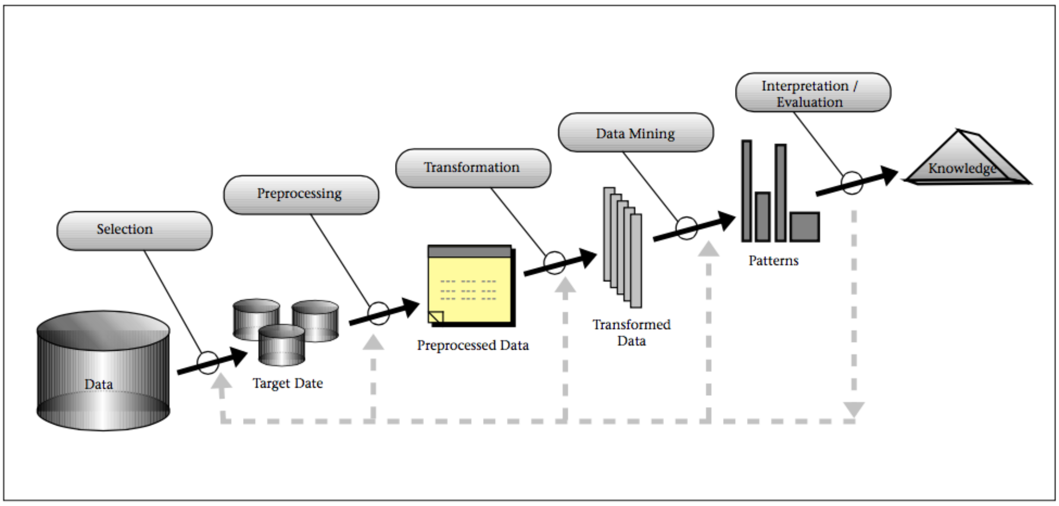


Figure 3.10: An overview of the steps of a KDD process (Fayyad et al. 1996b)

That can imply multiple iterations of the steps involving data preparation, knowledge evaluation, search for patterns, and refinement. According to Miller and Han (2003) and Maimon and Rokach (2006b) the KDD process can be subdivided into nine steps whereas data mining can be seen as the core process. The first step is to understand the application domain and define the goals. Second is the selection of relevant data or variables. The third step is about preprocessing and cleaning, such as handling missing values or outliers. Transformation is done in the fourth step, which includes dimension reduction or attribute transformation. The next three steps are all related to data mining. Namely the selection of an appropriate data mining task, and a matching algorithm as well as the employment of the algorithm. The eighth step is the evaluation and interpretation of the mined patterns. Using the discovered knowledge is part of the last step. That means, the knowledge can be incorporated into other systems which allows conclusions about the effectiveness of the KDD process (Maimon and Rokach 2006b, Fayyad et al. 1996b). The entire process involves iterations and can contain loops between any two steps.

The discovery goal of a KDD process can be further subdivided into prediction and description goals. Whereas the description branch focuses on presentation of patterns, and thus supports understanding (e.g. by visualizing techniques), the predictive part aims to build a behavioural model. Fayyad et al. (1996b) pointed out that the border between these groups is not clearly defined but is helpful in order to understand the discovery goal. Maimon and Rokach (2006a) provided a taxonomy of data mining methods in order to get a better understanding of the variety and usage of methods (see also figure 3.11).

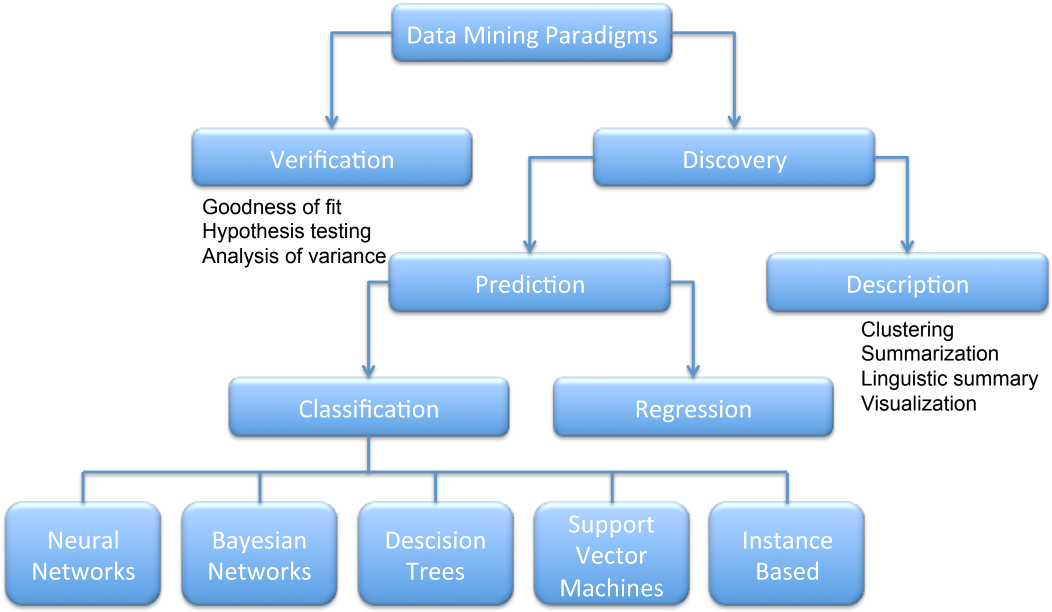


Figure 3.11: Data mining taxonomy (adopted from Maimon and Rokach 2006b)

A special case of KDD is the geographical knowledge discovery (GKD) (Miller and Han 2003). The vast amount of spatial and spatiotemporal data facilitates the understanding of more complex relationships, such as human-environment interaction. In order to extract spatial pattern from geographical data specialized tools are needed (Rinzivillo et al. 2008, Yuan et al. 2004).

> There is an urgent need for effective and efficient methods to extract unknown and unexpected information from datasets of unprecedentedly large size (e.g., millions of observations), high dimensionality (e.g., hundreds of variables), and complexity (e.g., heterogeneous data sources, space-time dynamics, multivariate connections, explicit and implicit spatial relations and interactions). (Guo and Mennis 2009, p. 403f.)

Additionally, in contrast to other information domains in a KDD process – which can be highly dimensional – geographic information has up to four dimensions which are interrelated (Miller and Han 2003). Spatial data has specific inherent properties such as topological relations or distances to other elements (Rinzivillo et al. 2008, Miller and Han 2003). Another characteristic was published by Tobler (1970) who suggests the following to be the first law of geography:

> [...] everything is related to everything else, but near things are more related than distant things. (Tobler 1970, p. 236)

Regarding the GKD process, it is important to factor in these spatial properties in order to derive new spatial patterns. This can be done by either making the spatial relationships explicit before or by using specialized algorithms. The advantage of the first approach is that known data mining techniques can be used. However, the latter approach allows the dynamic exploitation of spatial features (Rinzivillo et al. 2008). A lot of research is done to address the challenges related to GKD (e.g. Chawla 2005, Mennis and Liu 2005, Guo et al. 2003, Miller and Han 2003, Andrienko and Andrienko 1999, Openshaw et al. 1987). The presented methodology combines various methods in order to gain knowledge about spatial pattern.

As mentioned above (see chapter 2.2, predictive modelling approaches are similar or identical to settlement pattern analysis (e.g. fitting the findings to a theoretical distribution) as well as site catchment and territorial analysis (e.g. deriving cost surfaces). The idea of extrapolation in order to identify new, yet unknown, prehistoric settlements (or at least test hypothesis, assumptions, or scenarios of such) makes it a predictive modelling approach. Moreover, the usage of those terms is not always consistent. In the following, the terms are used as follows. Settlement pattern analysis focusses on the point pattern, thus identifying if a randomly dispersed, uniform or clustered pattern can be assumed. Site catchment and territorial analysis is about the determination of environmental influences and area(s) of influence. Predictive modelling tries to predict further suitable, and thus likely, habitats. Functional settlement pattern analysis serves as an umbrella term for all of the terms mentioned above.

3.2.1 Point Pattern Analysis

Excavation sites can be seen as points which are distributed within a region. This set of locations is called a spatial point pattern (Diggle 2014). When determining structured patterning in archaeological (point-like) distributions, point pattern analysis is used. Other possible spatial patterns are continuous or lattice processes (Bartlett 1974) but are not relevant in this case because the focus is on the locations of the excavation sites. In terms of point pattern analysis the observed points are often referred to as events to distinguish between observed points and other arbitrary points (Diggle 2014). If not specified otherwise, points and events are used synonymously in this thesis. A point process in this case is a stochastic model which focuses on the location of events in an area – at least when spatial point processes are analysed (Cressie 1993). When working with point processes it is not sufficient to only analyse the data itself. Information from non-data locations (often randomly dispersed points) should be considered to be able to compare and interpret the data (Badde-

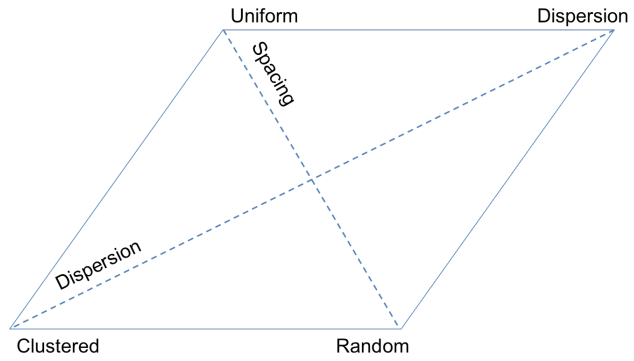


Figure 3.12: Describing point pattern (adapted from Kariel 1970)

ley 2010). Typical scientific questions are about intensity or interpoint interaction (see figure 3.12) and the effect of the surroundings (Baddeley 2010). Originally developed in the field of plant ecology and geography (e.g. Ripley 1987, Pielou 1977, Goodall 1970, 1952), the usage of methods to analyse the distribution of points in archaeology increased with a slight delay (Hodder and Orton 1976). Whereas the use of spatial point pattern analysis can be problematic in terms of intra site relations, the usage in between-site analysis is more valuable (Hodder and Orton 1976). Regarding spatial point pattern analysis, summary statistics are a simple and fast approach to get first insights into the pattern. This includes e.g. the average distance or the number of points in a certain area. Another common procedure, which requires more statistical theory, is to test on complete spatial randomness (CSR), namely to compare the observed pattern to a homogeneous Poisson process (HPP) (Diggle 2014, Baddeley 2010, Cressie 1993, Bartlett 1974). This means, by applying HPP – which is referred to as Poisson process as well – the probability of an event is not influenced by nearby events (Bivand et al. 2013). In the case of spatial analysis the pattern would be randomly dispersed if geography does not influence a process. O'Sullivan and Unwin (2010) described this as the

> ultimate null hypothesis for any geographer to suggest. (O'Sullivan and Unwin 2010, p. 98)

Numata (1961) published a dataset containing 65 locations of Japanese black pine saplings which might be regarded as random (see figure 3.13 a). In contrast, Ripley's (1977) (extracted from Strauss (1975)) dataset shows 62 redwood seedlings which appear to be clustered (see figure 3.13 b). In the case of redwood seedlings it is known that they agglomerate around redwood stumps, therefore an environmental influence can be assumed (Diggle 2014). A third example (see figure 3.13 c), also published by Ripley (1977), shows a more or less regularly dispersed pattern. It is important to point out that an identification of a pattern does not necessarily explain

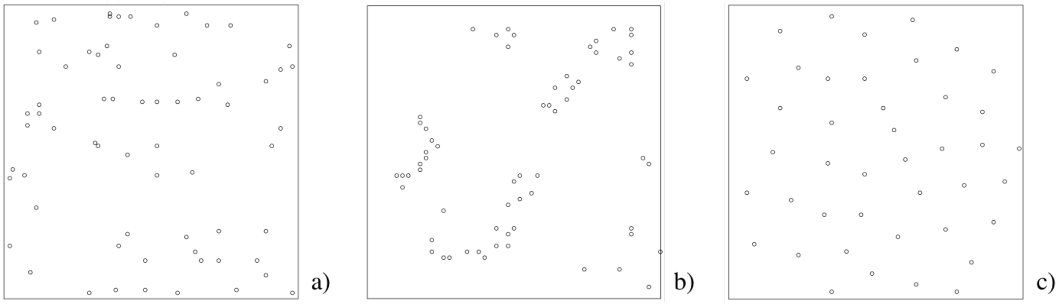


Figure 3.13: Location of a) 65 Japanese black spine saplings in a square of side-length 5.7 metres (Numata 1961), b) 62 redwood seedlings in a square of side-length 23 metres (Ripley 1977, Strauss 1975), and c) 42 cell centres in a unit square (Ripley 1977) (as cited in (Diggle 2014))

anything, but rather helps to interpret a spatial process which leads to that pattern (Stanislawski 1973).

> Our interest in CSR is that it represents an idealized standard which, if strictly unattainable in practice, may nevertheless be tenable as a convenient first approximation.(Diggle 2014, p. 10f)

CSR is helpful in terms of exploring a dataset and formulating hypotheses based on the resulting patterns. The rejection of CSR is the minimum requirement for further analysis (Diggle 2014). Baxter (2003) stated that a demonstration of a departure from complete spatial randomness does not explain the pattern. Davis (1973) showed that, depending on the used method(s), the result does not necessarily have to be unique. Thus, the distinction between the three pattern categories random, aggregated (or clustered) and regular is simplified but useful to start with (Diggle 2014). The careful selection of the study region is important in order to get satisfactory results. In some cases the region is objectively determined by the research subject such as settlements on an island. In other cases a – in some sense – representative subregion is selected from a larger region. Rather than analysing a single region, a subdivision into a large number of smaller regions, so called quadrats, can simplify the analysis. The term quadrat is derived by the Uppsala school of plant ecologists (Du Rietz 1929 (as cited in Diggle 2014), Greig-Smith 1952). Quadrats or rectangles are usually used to subdivide the observed region but it is theoretically possible to use other grids such as hexagons or triangles (Hodder and Orton 1976). The quadrat location can either be randomly or contiguously dispersed over the area and is constrained so that no overlapping is possible (Cressie 1993). Clarke (1946) used this method combined with the Poisson distribution to determine whether flying-bomb attacks in World War 2 were randomly dispersed or clustered (see table E.01). The examined region was subdivided into quadrats of

$\frac{1}{4}km^2$ each. The dispersion is random if every cell has an equal and independent probability of containing a point. The Poisson function is used to return the probability that a cell contains exactly x points (King 1968) and the Poisson distribution represents a randomly dispersed pattern (Diggle 2014).

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{3.1}$$

with:
$x =$ the number of points and
$\mu =$ average number of points per cell (sometimes also referred to as *lambda*). CSR for a spatial point pattern is given when the intensity (mean number of points per quadrat) follows the Poisson distribution and when the points are an independent random sample from the uniform distribution (Diggle 2014). To evaluate the predictive quality of a model, a goodness of fit test is used (Rönz and Strohe 1994). Clarke (1946) used a Pearson χ^2 test to compare the estimated Poisson distribution with the observed data. It tests whether an observed distribution is consistent with another distribution (null hypothesis) (Pearson 1900). Therefore a χ^2-value is calculated, which is compared to the associated critical value (the quantil of the χ^2 distribution which is determined by the degree of freedom and the level of significance).

$$\chi^2 = \sum \frac{(frequency_{observed} - frequency_{expected})^2}{frequency_{expected}} \tag{3.2}$$

Assuming a significance level of $\alpha = 0.05$ and the degree of freedom $df = 4$, the critical value is $\chi^2_{0.05,df=4} = 9.488$ (see table F.01 for the list of critical values). The significance level α determines the probability of so called type 1 errors, namely the probability to incorrectly reject the null hypothesis (Schlotzhauer 2007). The degrees of freedom specify the number of parameters which can be varied in a statistical calculation. In the case of a Pearson χ^2 test the degrees of freedom are defined as $n-1$ with $n =$ number of classifications. In the example of bomb hits, the null hypothesis H_0 means that the bombs were falling randomly (the Poisson distribution matches the empirical distribution) which can be rejected if $\chi^2 > \chi^2_{0.01,df=4}$ (Schönwiese 1992). The critical value in table F.01 leads to the equation $1.17 \not> 13.277$, therefore the null hypothesis cannot be rejected. For a meaningful interpretation of the result a p-value can be used which indicates the evidence against the null hypothesis. The smaller the p-value the stronger the evidence against H_0. The p-value can be calculated as follows:

$$p = 1 - F(x) \tag{3.3}$$

where F(x) correspond to the cumulative distribution function (CDF) which is defined as:

$$F(x) = \int_0^x \frac{x^{\frac{v}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} = \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \tag{3.4}$$

with:
$\Gamma(x) = (x-1)!$ for all real numbers.

Greig-Smith (1952) used the same approach and tested it against different sizes of quadrats and possible patterns. The cell size is important in order to identify the pattern and can lead to wrong conclusions if incorrectly chosen (Diggle 2014, 2003, Greig-Smith 1952) and therefore is a point of criticism. Other possibilities to optimize the grid number were published by Greig-Smith (1983), Moellering and Tobler (1972). The suggestion according to Greig-Smith (1952) is to test varying sizes and to include a relative variance test in addition to the χ^2 test. An overview of additional quadrat method indeces can be found in table E.02. The relative variance rel_{var} is defined as

$$rel_{var} = \frac{s^2}{\bar{x}} \tag{3.5}$$

with s^2 = variance and $\bar{x}$ = mean and was used by Clapham (1936) to compare the observed pattern with the Poisson distribution. The variance equals the mean in a Poisson distribution and therefore indicates a random pattern. If the null hypothesis can be rejected, the measurement of the so called departure from CSR is the next step (Cressie 1993). A clustered pattern can be assumed when the ratio is greater than 1, a more or less regularly dispersed pattern when the value is less than 1. Other points of criticism are the necessity of having a rectangular grid (Hodder and Orton 1976), or the alternation of results in a sequence of quadrat sizes (Pielou 1969), especially in terms of archaeological excavations (King (1968) and Stiteler and Patil (1971) (as cited in Hodder and Orton 1976)) point out that the shape of a quadrat itself can be misleading.

Diggle (2014) stated that the quadrat sampling method still remains popular but can be impractical. That leads to the development of a number of distance methods focussing on the distance to adjacent features, which are more appropriate for most archaeological studies (Hodder and Orton 1976). A common approach was presented by Clark and Evans (1954) who compared the ratio between the expected and the observed mean nearest neighbour distances. In the beginning, the distance to the nearest neighbour for each point has to be measured. The sum of distance values r is divided by the total number of observed points n to get the mean nearest neighbour distance $\bar{r}_o$.

$$\bar{r}_o = \frac{\sum r}{n} \tag{3.6}$$

The expected mean nearest neighbour distance $\bar{r}_e$ of a random distribution of points is only dependent on the density ρ of the point and can be defined as

$$\bar{r}_e = \frac{1}{2\sqrt{\rho}} \tag{3.7}$$

where ρ is calculated as follows:

$$\rho = \frac{n-1}{A} \tag{3.8}$$

with $A =$ size of the region. The distribution is random if the ratio $R = 1$, a value less than 1 indicates an aggregated, a value greater than 1 a regularly dispersed pattern. Again, a goodness of fit test can be used to test if the observed distribution differs significantly (Clark and Evans 1954, Hodder and Orton 1976).

$$R = \frac{\bar{r}_o}{\bar{r}_e} \tag{3.9}$$

This distance method can be extended in order to include nth nearest neighbour (Greig-Smith 1983, King 1969) and therefore reduces the risk of falsely identifying clustered patterns. Another modification was given by Washburn (1974) who suggested the use of the median nearest neighbour distance $\bar{r}_m$ rather than the mean, especially when only few very large distances are known. Several other distance methods were published in order to fit special research or data source related requirements (e.g. Getis and Franklin 1987, Holgate 1965, Catana 1963, Pielou 1959, Cottam and Curtis 1949). The selection of the investigated area may influence the result of nearest neighbour methods, because the boundaries are often subjectively determined (Kariel 1970, Getis 1964). In the case of Clark and Evans's 1954 R value, the area is needed to calculate the density ρ and therefore has influence on the expected mean nearest neighbour distance $\bar{r}_e$. The need of a boundary may lead to errors due to lack of data at the border area. A potential nearest neighbour may occur just outside the chosen boundary and is therefore not factored in the calculation (Diggle 2014, 2003, Hodder 1971, Clark and Evans 1954). To minimize this boundary effect, a buffer zone can be used (Hodder and Hassall 1971), points are only used if the nearest neighbour is closer than the boundary (Dacey 1963) or another weight is assigned to points close to the border (Getis and Franklin 1987).

Both the quadrat method as well as the distance method are helpful to identify the intensity of a spatial point pattern. The presented methods are good to get a first impression of the empirical data. In order to get a more precise overview of the spatial variations, several methods are available to analyse the observed distribution on CSR in more detail. The most common functions for stationary point processes are presented in the following. Namely the

- G-function (nearest neighbour distribution function),
- F-function (spherical contact distribution function),
- K-function (Ripley's K-Function).

In contrast to the summary statistics described above, these methods can be used to identify certain thresholds. One example is given by Diggle (2014) who stated that the G-function can be used to objectively specify the minimum area of trees needed. The author argues that trees usually need sunlight and nutrients and therefore a certain area for their roots and crowns is required. By applying the G-function the small inter-event distances can be identified and used as a threshold.

The nearest neighbour distribution function (G) is defined as the probability distribution of the distance of one point p_i of a point process X and the nearest neighbour p_{nn} and $p_i, p_{nn} \in X$. Thus, this function (also called event-to-event or inter-event distribution) describes the probability of other events occurring within a certain distance of an event (Diggle 2014, 2003, Baddeley 1998, Cressie 1993). The estimated distribution function (EDF) can be described as follows:

$$\hat{G}(r) \equiv \frac{\sum_{i=1}^{n} I(r_{i,A} \leq r, d_i > r)}{\sum_{i=1}^{n} I(d_i > r)}, r > 0 \tag{3.10}$$

with:
I being the indicator function which is either 1 (true) or 0 (false),
$r_{i,A}$ indicating the distance from a point p_i to the nearest neighbour p_{nn} – the nearest neighbour distance,
d_i meaning the distance from point p_i to the nearest boundary, and
r the observed distance (Cressie 1993)
which can be reduced to

$$\hat{G}(r) \equiv n^{-1} * I(r_{i,A} \leq r) \tag{3.11}$$

with:
$n =$ the total number of points $\in X$

if a boundary effect can be eliminated (Diggle 2014). The distribution of the nearest neighbour distances of a homogeneous Poisson process cannot be expressed in closed form due to the edge effects. The approximate distribution function under CSR is

$$G(r) = 1 - exp(-\lambda \pi r^2), r \geq 0 \tag{3.12}$$

where λ = *intensity* (expected number of points). The empty space function for the Poisson process is identical to the equation above. An easy way to compare the two distributions is by simply plotting them (e.g. see figure 3.14 a). Rather than using the approximate distribution function, a sample mean $\hat{G}(r)$ can be computed using randomly dispersed points. If the line of the EDF $\hat{G}(r)$ equals the line of the approximate distribution function $G(r)$, a randomly dispersed pattern can be assumed. The observed locations are assumed to be closer to each other with a clustered pattern, thus values $\hat{G}(r) > G(r)$ indicate a clustered, $\hat{G}(r) < G(r)$ a regularly dispersed pattern (Baddeley 2010) (see also plots on the left in figures 3.14 - 3.16).

The G-function is often used in combination with the F-function. While the G-function measures the distance between events, the F-function measures the distance between a random point rp and an event p_i where $rp \notin X$. This means a point in this case is not equal to an event but a randomly dispersed point which is not in X. Thus it is the probability distribution of the distance between a random point rp and its nearest neighbour event p_i. The EDF $\hat{F}(r)$ looks similar to the G-function. In contrast to the G-function the F-function measures the empty spaces in the area A which can be useful in exploratory analysis to determine the sizes of the gaps in the point process.

$$\hat{F}(r) \equiv m^{-1} * I(r_{i,R} \leq r) \tag{3.13}$$

with:
m meaning the number of randomly dispersed points, and
$r_{i,R}$ indicating the distance from a random point rp to the nearest neighbour event p_i.

The same approximate distribution function that was used for $G(r)$ can be used to derive $F(r)$, whereas the interpretation of the F-function is reverse. Clustered patterns are expected to be further away from random points than randomly dispersed points. This means that $\hat{F}(r) > F(r)$ indicates a regularly dispersed pattern, $\hat{F}(r) < F(r)$ a clustered pattern (see also plots in the middle of figures 3.14 - 3.16).

The third function is the Ripley's K-function (sometimes also called reduced second moment function). It measures the number of events within a specified distance and describes the point process at various scales (Dixon 2002, Ripley 1977) which is the main difference between the K-function and the G-, respectively F-function. A point process can show a combination of patterns according to the scale (e.g. trees can be regularly dispersed at small scale but may be clustered in large scale). The K-function allows to identify these characteristics.

$$K(r) \equiv \lambda^{-1} E(n_{extra}), r \geq 0 \tag{3.14}$$

where n_{extra} = *number of extra events within distance r of an arbitrary event*. In other words the K-function means that λ *K(r)* equals the expected number of additional events within the distance r. Let s_1,...,s_N be all events in the study region so that $s_1,...,s_N \in X$ then the EDF $\hat{K}(r)$ can be defined as follows:

$$\hat{K}(r) = \hat{\lambda}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} I(||s_i - s_j|| \leq r)/n, r \geq 0, i \neq j \tag{3.15}$$

Several other EDFs were published in order to factor in the edge effects (Baddeley 1998, Ripley 1988, Ohser 1983, Ripley 1976). The true value for a homogeneous Poisson process is

$$K(r) = \pi r^2 \tag{3.16}$$

which is used to compare it with $\hat{K}(r)$ for inferential purposes. Similar to the G-function values close to $K(r)$ indicate a random pattern whereas $\hat{K}(r) < K(r)$ indicate a regularly dispersed distribution and $\hat{K}(r) > K(r)$ a clustered pattern (see also plots on the right in figures 3.14 - 3.16).

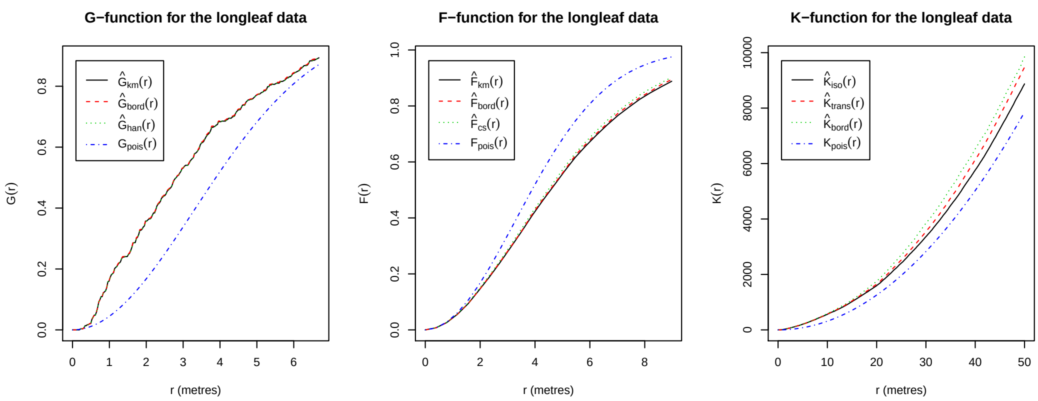


Figure 3.14: From left to right: Applied G-function, F-function and K-function to the longleaf dataset. The blue dashed line indicates the Poisson distribution whereas the other lines symbolize varying EDFs.

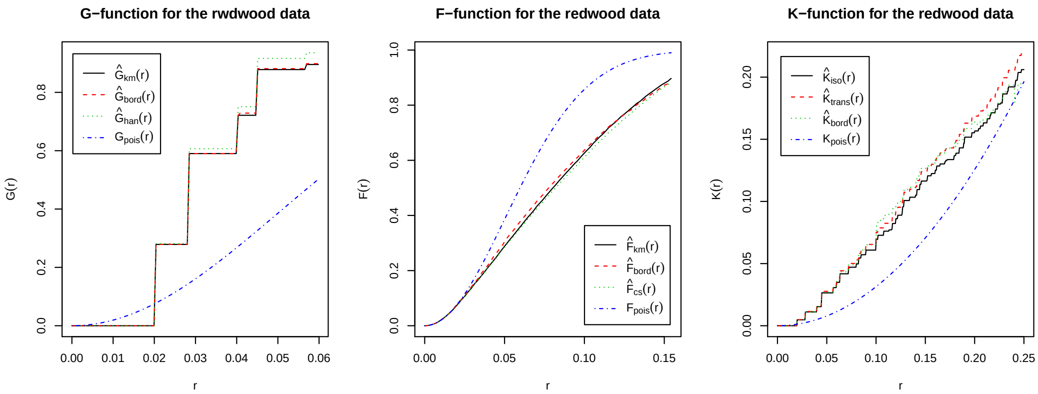


Figure 3.15: From left to right: Applied G-function, F-function and K-function to the redwood dataset. The blue dashed line indicates the Poisson distribution whereas the other lines symbolize varying EDFs.

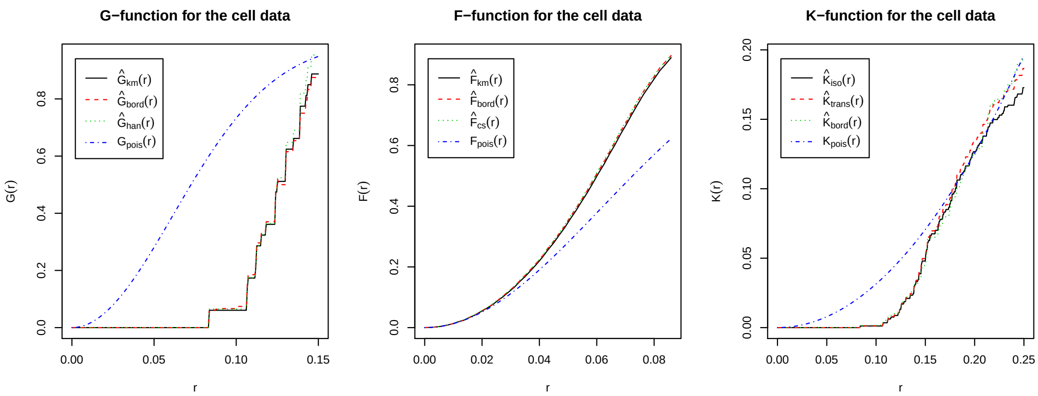


Figure 3.16: From left to right: Applied G-function, F-function and K-function to the cells dataset. The blue dashed line indicates the Poisson distribution whereas the other lines symbolize varying EDFs.

As shown in figures 3.14 - 3.16, the functions give a good impression of the observed point process. Random patterns have a high variability and observed random processes will never match the Poisson Distribution perfectly. A common procedure to make this result statistically reliable is to perform a Monte Carlo simulation. Barnard (1963) pointed out that even with simple case studies a test of significance of the model – the use of Monte Carlo simulations – is reasonable. The aim of a Monte Carlo simulation is to simplify a hard combinatorial problem by selecting a statistical sample in order to approximate the initial problem (Andrieu et al. 2003). The idea is to rank the observed data value against a set of randomly generated values and thus determining the significance level of the selected test statistic (Besag and Diggle 1977). This is possible because the principle of Monte Carlo simulations is based on the law of great numbers (Graham and Talay 2013). By performing s simulations from the null hypothesis, the variation inherent in random processes can be depicted (Baddeley 2010). Metropolis and Ulam (1949) introduced the Monte Carlo simulation and illustrated its advantages by the example of the game solitaire.

> To calculate the probability of a successful outcome of a game of solitaire (we understand here only such games where skills plays no role) is a completely intractable task. [...] Obviously the practical procedure is to produce a large number of examples of any given game and then to examine the relative proportion of success. (Metropolis and Ulam 1949, p. 336)

The authors point out that the result of a Monte Carlo simulation only allows conclusions about the probability of an occurrence. Therefore the estimation is not an exact certainty, but rather a significance measure if the number of trials is great enough. Hope (1968) tested various examples of simulations in order to show the validity of the Monte Carlo approach. The author concludes that $s = 100$ is adequate for a significance level of 5%. Monte Carlo simulations are particularly applicable when data is sparse (Besag and Diggle 1977) or the calculation is unsolvable (Metropolis and Ulam 1949). Ripley (1981), Besag and Diggle (1977), Ripley (1977) and Besag and Clifford (1989) adapted the methods for spatial statistics. Cliff and Ord (1981, 1973), **?** used this approach in order to analyse spatial autocorrelation and its validity for small sample sizes. Besag and Diggle (1977) presented several use cases of Monte Carlo simulations for spatial patterns, such as pattern similarity, space-time interaction or – which is important in this thesis – spatial point pattern. The Monte Carlo simulation – as the test of CSR in general – is

> ...rarely to be treated as an end in itself, its purpose being more usually as an aid in suggesting further hypothesis and relevant data collection. (Besag and Diggle 1977, p. 327)

Generally spoken, if u_1 is an observed value of U and $u_i : i = 2, ..., s$ are randomly selected values of U and $u_{(j)}$ represents the jth largest value of $u_i : i = 1, ..., s$ – note that here, u_i includes u_1 – the probability that u_i equals $u_{(j)}$ is s^{-1}.

$$P\{u_1 = u_{(j)}\} = s^{-1}, j = 1, ..., s \tag{3.17}$$

Therefore the probability if u_1 ranks kth largest or higher can be exactly determined by $\frac{k}{s}$. That value can be used to reject or accept a null hypothesis H_0 by comparing it to the observed rank (Diggle 2014). The equation 3.17 is only valid, if the values u_i are all unique and it is only a one-sided test. In terms of CSR the test is usually a two-sided test which is testing whether the values range within an envelope,

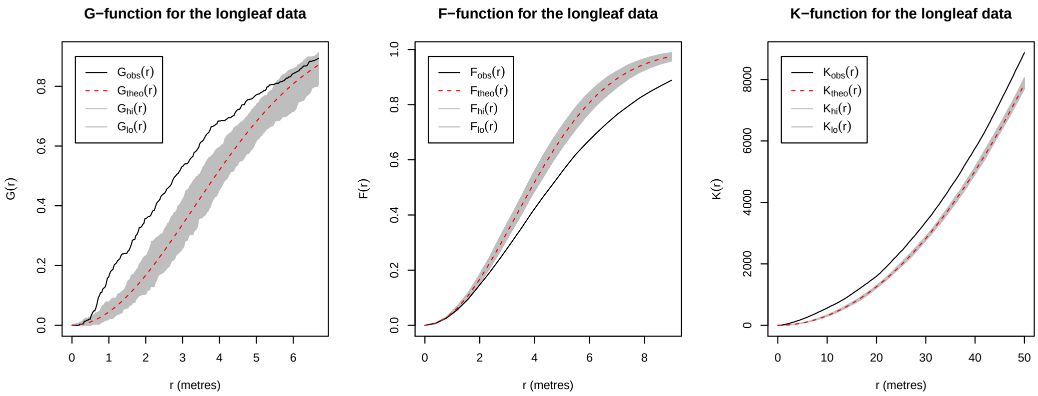


Figure 3.17: From left to right: Applied G-function, F-function and K-function to the longleaf dataset. The red dashed line indicates the theoretical, the black line the observed distribution, and the grey area indicates the simulated envelope.

formed by maximum and minimum values, depending on the level of significance (Baddeley 2010). These boundaries can formally be described as follows:

$$L(u_i) = min(F(u_i)) \tag{3.18}$$

as lower boundary, and

$$U(u_i) = max(F(u_i)) \tag{3.19}$$

as upper boundary where $F(u_i)$ describes a function such as the G-, F-, or K-function. If the values are outside of the envelope defined by $L(u_i)$ and $U(u_i)$ the null hypothesis needs to be rejected. A more detailed information about the point process is provided if applied to the datasets above (as can be seen in figures 3.17 - 3.19). The black line $< function_letter >_{obs}$ (r) equals the variable $< \hat{function_letter} > (r)$. The red line indicates the theoretical distribution and the grey area indicates the simulated envelope. The shown tests are necessary in order to get a first impression of the pattern. A clustered pattern may not be surprising because human movement behaviour is expected to be constrained and determined by certain parameters (Hodder 1977).

3.2.2 Maximum Entropy Modelling

One fundamental step in cluster analysis is the selection of variables which determine the clustering result (e.g. Ketchen and Shook 1996). Clustering methods do not rescale the input variables (like PCA does), in order to minimize correlation effects. Ketchen and Shook (1996) distinguishes three different ways to determine the input

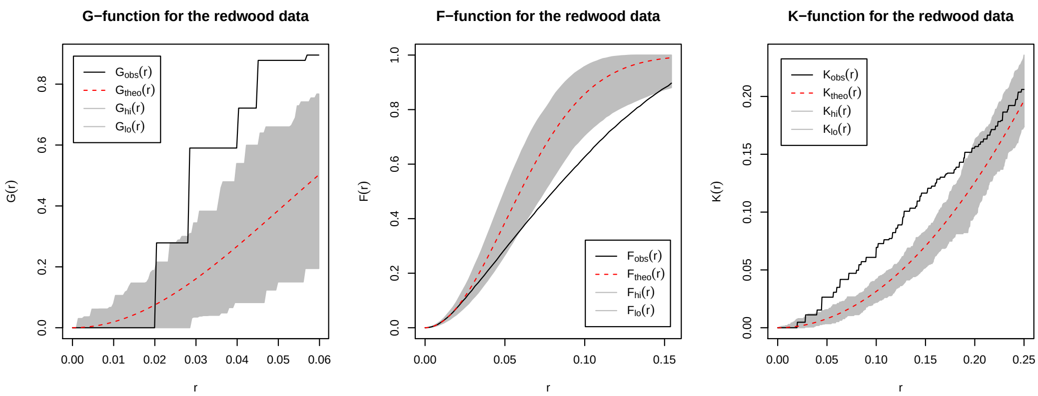


Figure 3.18: From left to right: Applied G-function, F-function and K-function to the redwood dataset. The red dashed line indicates the theoretical, the black line the observed distribution, and the grey area indicates the simulated envelope.

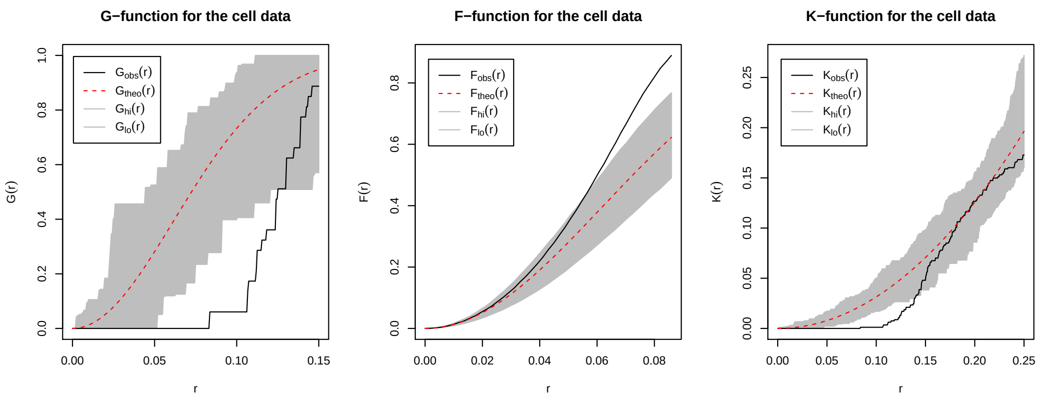


Figure 3.19: From left to right: Applied G-function, F-function and K-function to the cells dataset. The red dashed line indicates the theoretical, the black line the observed distribution, and the grey area indicates the simulated envelope.

variables. One possible approach would be to consider as many variables as possible (inductive approach), because it is not known in advance which variables differentiate among the items (Ketchen and Shook 1996, McKelvey 1978, 1975). Hambrick (1983) used this approach in order to develop a taxonomy of mature industries. The author used ten environmental variables whereas no a priori expectations about the likely nature of the resultant types were made. When the number and suitability of variables are strongly tied to theory, a deductive approach is used (Ketchen and Shook 1996, Ketchen et al. 1993). Punj and Stewart (1983) suggest to use a deductive approach because cluster algorithms derive the most internally consistent groups across all variables and thus can cause a deterioration of a solution's validity (as mentioned in Ketchen and Shook 1996). Similar to the inductive approach the cognitive approach avoids to make theory-based predictions. In contrast to the inductive approach, the cognitive approach considers expert knowledge and relies on perceptions of expert information (Ketchen and Shook 1996). As mentioned above (see chapter 2.2.3) this thesis aims for a data driven approach, therefore the inductive way is used. To avoid deterioration of the clustering result the influences of the variables is determined. Only relevant variables are used in the next step (the clustering). To achieve that, a MEM is used. One reason for choosing this model is that no dependent variable is needed, which does not exist in this case. Due to the inductive approach, no assumptions about dependencies or independencies are made and only the location serves as input and is tested against other (e.g. environmental) variables. The MEM tests, whether the variables are important predictors or not (McMichael et al. 2014). Another advantage is the ability to use presence-only data rather than presence/absence data (Elith et al. 2011). In the case of spatial archaeology, presence data would indicate known former settlements. In contrast, absence data would be the information about locations where definitely no settlement existed. Due to the lack of absence data in many datasets and in order to provide a method which can easily be applied to other study areas presence-only data serves as input data. Another advantage is the ability to provide reasonable results even when dealing with small sample sizes or biased samples (McMichael et al. 2014, Phillips et al. 2006). Additionally it can utilize continuous as well as categorical data (Phillips et al. 2006). In this thesis the software MaxEnt (Version 3.3, https://www.cs.princeton.edu/ schapire/maxent/) is used which was developed for species distribution analysis.

Whereas the concept of maximum entropy is very old (Berger et al. 1996 stated, that it can be traced back to Biblical times) it is only applied to real world problems since computers are powerful enough (Berger et al. 1996). Jaynes (1957), a more recent pioneer of the MEM stated that the maximum entropy estimate is based on partial knowledge.

> ...the fact that a certain probability distribution maximizes entropy subject to certain constraints representing our incomplete information, is the fundamental property which justifies use of that distribution for inference; it agrees with everything that is known, but carefully avoids assuming anything that is not known. (Jaynes 1991, p. 1)

It is a commonly used predictive modelling technique in machine learning (especially in natural language processing), image reconstruction, signal processing or statistical physics and is one of the most popular tools in species distribution modelling approaches (Merow et al. 2013, Elith et al. 2011, Franklin 2009, Phillips and Dudík 2008, Elith et al. 2006, Phillips et al. 2006, 2004, Berger et al. 1996).

Maximum Entropy Modelling aims for estimating a target probability distribution by identifying the probability distribution of maximum entropy (Phillips et al. 2006). Berger et al. (1996) explained the method using the example of automatically translating a word from English to French. The author aims to model the decisions of an expert translator concerning the English word *in*. Lets assume, that the translator always used one of the following five french words (*dans,en,à,au cours de,pendant*). The probability model π assigns an estimate $\pi(x)$ for each word x in a finite set X (in terms of spatial analysis x is usually a finite number of points in a study area) so that it is constrained as follows:

$$\pi(dans) + \pi(en) + \pi(\text{à}) + \pi(\text{au cours de}) + \pi(pendant) = 1 \tag{3.20}$$

This equation has an infinite number of models π, so how can either of these distributions be justified? The idea of maximum entropy modelling is to identify the most uniform π satisfying the given constrains. To stick with the example given above lets further assume that either *dans* or *en* are chosen 30% of the time.

$$\begin{aligned} &\pi(dans) + \pi(en) = 3/10 \\ &\pi(dans) + \pi(en) + \pi(\text{à}) + \pi(\text{au cours de}) + \pi(pendant) = 1 \end{aligned} \tag{3.21}$$

The most uniform (most evenly allocation of probabilities) model π is achieved with the following assignments:

$$\begin{aligned}
&\pi(dans) = 3/20 \\
&\pi(en) = 3/20 \\
&\pi(\text{à}) = 7/30 \\
&\pi(\text{au cours de}) = 7/30 \\
&\pi(pendant) = 7/30
\end{aligned} \tag{3.22}$$

Lets further assume that the expert chooses either *dans* or *à* in half the cases. This complicates the identification of the most uniform π which satisfies these constraints.

$$\begin{aligned}
&\pi(dans) + \pi(en) = 3/10 \\
&\pi(dans) + \pi(en) + \pi(\text{à}) + \pi(\text{au cours de}) + \pi(pendant) = 1 \\
&\pi(dans) + \pi() = 1/2
\end{aligned} \tag{3.23}$$

That leads to the main question, what is the best approximation when modelling an unknown probability (Phillips et al. 2006)? Jaynes (1957) stated that the best distribution is the one which satisfies any constraints that one is aware of and has the maximum entropy. The entropy $H(\hat{\pi})$ (where $\hat{\pi}$ is the approximation of π) is defined as follows:

$$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x) ln \hat{\pi}(x) \tag{3.24}$$

The resulting entropy is non-negative and its maximum is the natural logarithm of all $x \in X$. Shannon (1948) stated that entropy measures the choice which is involved in the selection of an event. The more choices the higher the entropy of the involved distribution (which means that it is less constrained). Therefore a maximum entropy is achieved by applying the approximate probability distribution $\hat{\pi}$ with no unfounded constraints (Elith et al. 2011, Phillips et al. 2006). A constraint is defined by k features f and can be defined as follows:

$$f(j) = \begin{cases} 1 \text{ tranlated to } \textit{en} \text{ and } \textit{April} \text{ follows } \textit{in} \\ 0 \text{ otherwise} \end{cases} \tag{3.25}$$

where $1 \leq j \leq k$

$$\hat{\pi}(f) = \pi(f) \tag{3.26}$$

where $\hat{\pi}(f)$ describes the expected probability value whereas $\pi(f)$ describes the observed probability value with regard to the constraint. The equation 3.26 determ-

	Primal	Dual
problem	$\text{argmax}_{p \in P} H(p)$	$\text{argmax}_\lambda \Psi(\lambda)$
description	maximum entropy	maximum likelihood
type of search	constrained optimization	unconstrained optimization
search domain	$p \in P$	real-valued verctors $\{\lambda_1,\lambda_2,...\}$
solution	p_*	λ^*
Kuhn-Tucker theorem: $p_* = P_{\lambda*}$		

Table 3.2: The duality of maximum entropy and maximum likelihood (adapted from Berger et al. 1996)

ines that the constraints of the expected distribution need to match the observed distribution. In terms of maximum entropy it is recommended to use p^* which is defined as follows (e.g. Ratnaparkhi 1998, 1997, Berger et al. 1996):

$$\begin{aligned} p^* &= argmax_{p \in P}(H(\hat{\pi})) \\ P &= \{\hat{\pi} | \hat{\pi}(f) = \pi(f), j = \{1...k\}\} \end{aligned} \tag{3.27}$$

This constrained optimization problem find $p^* \in P$ which maximizes $H(\hat{\pi})$ can be solved using the mathematical theory of convex duality (Della Pietra et al. 1997). The main idea is that $\hat{\pi}$ can alternatively be described, considering all probability distributions of the form

$$q_\lambda(x) = \frac{e^{\lambda * f(x)}}{Z_\lambda} \tag{3.28}$$

with:
λ as vector of n real-valued feature weights (on other words a Lagrange multiplier λ_i is assigned to each feature f_i) ,
f as vector of all k features
Z_λ as normalizing constant so that q_λ sums to 1 (Phillips et al. 2006).

The convex duality shows that these so called Gibbs distributions (which maximize the likelihood) are equal to the maximum entropy probability distribution $\hat{\pi}$ (see also table 3.2).

> This duality is appealing, since p^* as a maximum likelihood model, will fit the data as closely as possible, while as a MEM, will not assume facts beyond those in the constraints. (Ratnaparkhi 1997, p. 8)

Therefore the MEM is solved by determining the distribution q_λ which maximizes the likelihood of the sample points (Berger et al. 1996). To calculate the optimal parameter values λ_i for each feature f_i, an iterative scaling algorithm (of Darroch

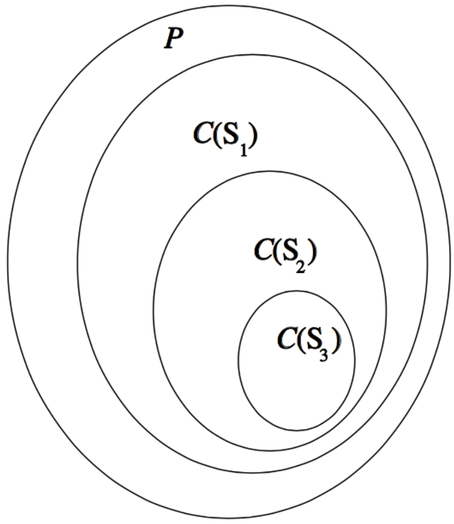


Figure 3.20: Decreasing number of possible models p^* by using a nested sequence of subsets $C(S_1) \supset C(S_2) \supset C(S_3)$ which delimits the initial set P corresponding to an increasing set $S_1 \subset S_2 \subset S_3$ (Berger et al. 1996)

and Ratcliff (Darroch and Ratcliff 1972)) is used which is tailored to the needs of the maximum entropy problem. In order to determine the constraining features f_i a subset S containing all active features is introduced. The collection of potential candidate features can be large and does not need to be relevant. S is defined as an empty set $\emptyset$ which is subsequently filled in an iterative process. Each step selects the f_i which maximizes the gain in the log-likelihood of the training data and is added to S afterwards. Thus, the size of S increases with every step whereas the number of potential models p^* decreases (see also figure 3.20). When MEM is applied to spatial data, the landscape of interest L can be described as background containing all possible location $l \in L$. The features f (and thus the constraints) can be defined as the probability density across L. Let z denote a vector of (environmental) parameters and $y = 1$ indicates presence, $y = 0$ indicates absence, than $f_1(z)$ describes the probability of (environmental) conditions across L. In order to derive the feature functions, either linear or nonlinear (e.g. hinge or threshold values) fitted functions are possible (see table H.01 for a detailed description of available fitting functions in MaxEnt). The result is a model which maximizes entropy and additionally assigns a variable importance to each input parameter (as shown in figure 3.21). This allows the use of as many input variables as possible in the beginning but minimizes the deterioration of the following cluster analysis.

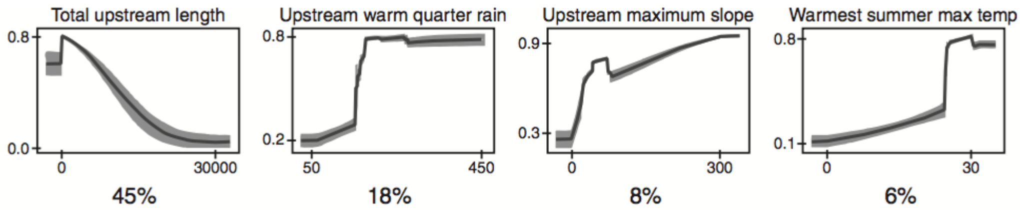


Figure 3.21: Partial dependence plots showing the marginal response to four variables (with the assigned variable importance below each graph) (Elith et al. 2011)

3.2.3 (Spatial) Clustering

> Do not forget that clusters are, in large part, on the eye of the beholder. (Estivill-Castro 2002, p. 65)

It is a common assumption among archaeologist that settlements of the same culture serve different functionalities.

> It is a common observation that there are fewer larger places than smaller ones in a region and that the larger centres provide a greater number and variety of goods than the small places do (Garner 1967, p. 322)

Observations of Bronze Age settlements underline this assumption. Bronze Age hillforts are supposed to represent the highest hierarchy of settlement types for that time (Stančič and Kvamme 1999). Dalton (1969) presented a model of socio-economic transactions which is based on the hierarchical idea of settlement concepts (see also figure 3.22). Struever and Houart (1972) tried to derive a multi-scaled interaction between Illinois Middle Woodland sites. Different settlement types were built which represent the interaction range (from intra-local to inter-regional interaction) as can be seen in figure 3.23. Another example was provided by Chang (1972) (see also figure 2.4) and chapter 2.2.1) who showed, that the hierarchical structure and functionalities can be very complex. The Central Place Theory (Christaller 1933) and other theories which seek to explain human settlement patterns also expect various functionalities and settlement types. It is for this reason that one premise of this thesis is the existence of at least two different settlement types. Other studies made use of different building types (e.g. hillforts) in order to distinguish between settlement functions, but this information is not necessarily given in an archaeological record. As mentioned in chapter 2.2, the location of a settlement itself is an important information.

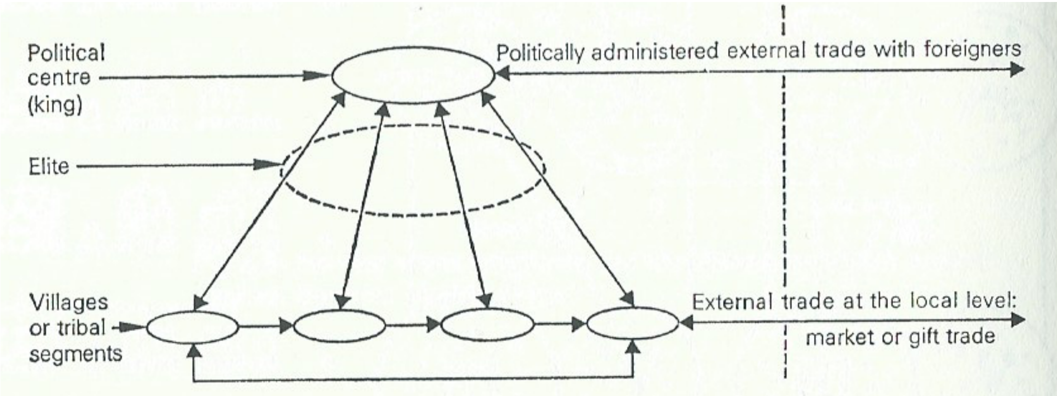


Figure 3.22: Socio-economic transactions in the primitive economy within a centralised political system (Dalton 1969)

> Sites in unexpected locations, or having distinctive archaeological features and catchments compared to other sites in similar locations, might indicate use for purposes other than subsistence – for defence, for the procurement of valuable raw materials, for the control of trade routes or markets, for social aggregation or for ceremonial and ritual. (Renfrew and Bahn 2005, p. 173f.)

Thus, another approach needs to be developed, which allows the identification of settlement types (or at least the distinction of such). The cluster analysis is a promising approach in order to find similarities in a given dataset and group similar objects, whereby each object is assigned to exactly one group. In this case, the environmental properties of the locations are compared in the cluster analysis and locations similar environmental conditions are grouped together. If the methods described in the chapter above (chapter 3.2.1) indicate a clustered pattern, further analysis can be done in order to gain detailed information about the clusters. The clustering process is unsupervised which makes it a commonly used technique for data mining approaches (Han et al. 2001). The aim is to group objects into clusters, so that the properties of objects within one cluster have a high similarity whereas the objects in other clusters are dissimilar (Bahrenberg et al. 2003, Han et al. 2001). Thus, a cluster analysis would allow to identify the different settlement functionalities based on the given input variables. Each entity (in this case settlement) is assigned to exactly one cluster (Bahrenberg et al. 2003). Cluster analysis differs from classification analysis which uses a predefined number of groups (Rencher and Christensen 2012). The use of thresholds is one possibility in classification analysis to subdivide the objects into different classes. Cluster analysis is an inductive approach which searches for similar properties of the (multivariate) observations without knowing the number of groups in advance (Rencher and Christensen 2012,

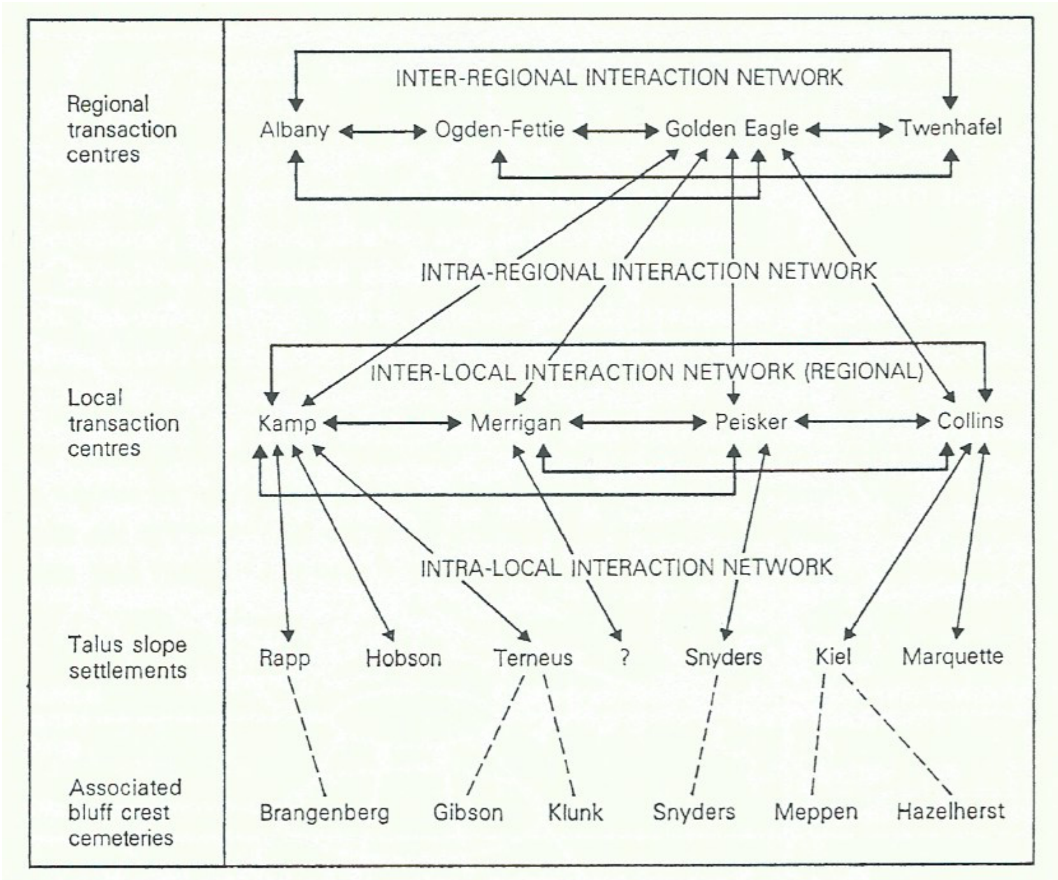


Figure 3.23: A hierarchy of interaction networks illustrated by specific Illinois Middle Woodland sites believed to exemplify the various settlement types involved (Struever and Houart 1972)

Bahrenberg et al. 2003). With regard to excavation sites, this concept can be used to identify groups of former settlements with similar (environmental) surroundings. The input data matrix can be written as follows:

$$Y = \begin{pmatrix} y_1^1 & y_2^1 & y_3^1 & . & . & y_i^1 & . & . & y_n^1 \\ y_1^2 & y_2^2 & y_3^2 & . & . & y_i^2 & . & . & y_n^2 \\ y_1^3 & y_2^3 & y_3^3 & . & . & y_i^3 & . & . & y_n^3 \\ . & & & & & & & & \\ . & & & & & & & & \\ y_1^j & y_2^j & y_3^j & . & . & y_i^j & . & . & y_n^j \\ . & & & & & & & & \\ . & & & & & & & & \\ y_1^p & y_2^p & y_3^p & . & . & y_i^p & . & . & y_n^p \end{pmatrix}, \tag{3.29}$$

where y_i is an observation and y^j the used variable (e.g. the distance to something). The aim is to cluster n rows into g clusters whereas g is not previously known (Rencher and Christensen 2012). A high number of different clustering methods exists and cluster analysis functions as an umbrella term for all the different approaches (Estivill-Castro and Yang 2000). The various algorithms differ in the sensitivity to small perturbations, the sensitiveness towards the order of the data, and the definition of similarity between clusters (Meilă 2007, Wagner and Wagner 2007, Bahrenberg et al. 2003). The similarity of objects (the cluster model) is often based on some kind of distance measure – which is not necessarily an euclidean distance measure. Other clustering techniques use cluster centroids as a basis – and try to minimize the mean distance of the objects to the cluster center –, compare the within and between-cluster variability or use the correlation of variables (Rencher and Christensen 2012). Several factors should be considered when it comes to choosing the cluster algorithm, such as the application goal, the characteristic of the data, or the tradeoff between quality and speed (Han et al. 2001). Two common approaches are the hierarchical clustering and partitioning (Rencher and Christensen 2012, Manly 2004), Han et al. (2001) additionally mentioned density based methods, grid based methods, and constrained based cluster analysis. Hierarchical clustering can be subdivided into bottom-up (sometimes also referred to as agglomerative) and top-down (sometimes also referred to as divisive) methods which differ in the starting situation. It either starts with n clusters (meaning, each observation is a cluster of its own) and ends up with all entities being in one cluster (bottom-up), or all objects initially belong to one cluster and are further divided (top-down) (Rokach and Maimon 2005). A similarity measure is needed in order to detect the most similar clusters (thus the two clusters which are merged to one cluster in a next step). Three different measures are common, namely the single linkage, the complete linkage, and the average linkage clustering (Rokach and Maimon 2005, Bahrenberg et al. 2003, Jain et al. 1999). The single linkage clustering uses a pairwise comparison of objects and uses the minimum distance of any object of one cluster and any object of the other cluster to determine the distance of clusters.

$$d_{Cr,Cs} = Min\{d_{jk}, j \in Cr, k \in Cs\} \tag{3.30}$$

with:
Cr and Cs as two clusters
$d_{Cr,Cs}$ as the distance between the two clusters

The complete linkage clustering is similar to the single linkage clustering but uses the maximum instead of the minimum distance between the objects.

$$d_{Cr,Cs} = Max\{d_{jk}, j \in Cr, k \in Cs\} \tag{3.31}$$

The average linkage clustering uses the average distance between all objects of one cluster and all objects of the other cluster (Rokach and Maimon 2005, Bahrenberg et al. 2003).

$$d_{Cr,Cs} = \frac{1}{n_r} * \frac{1}{n_s} \sum_{j \in Cr} \sum_{k \in Cs} d_{jk} \tag{3.32}$$

The single linkage clustering is the so called chaining effect, which results in few relatively large (in comparison to the total number of objects) clusters. The average linkage clustering tends to split elongated clusters and merge neighbouring elongated clusters (Rokach and Maimon 2005). More compact clusters are derived when the complete linkage clustering is used. Rencher and Christensen (2012) also mentions other approaches such as the centroid clustering, median clustering, Ward's method, and flexible beta method. Typically, the result is a dendrogram (see figure 3.24) where the built clusters and its containing objects can be seen for each step (Rencher and Christensen 2012, Manly 2004, Han et al. 2001, Tryfos 1998). By cutting the dendrogram at the desired similarity level, the clustering result is obtained. The advantage of the hierarchical clustering approach is the result is not one partition of g clusters but multiple nested partitions. Disadvantaged are the computing costs and the disability of back-tracking (Rokach and Maimon 2005). In contrast, partitioning arbitrarily chooses g centres or starts with an initial partitioning and reallocates the observations based on some optimality criterion Rencher and Christensen (2012). The most popular clustering algorithm which is based on partitioning is the k-means clustering algorithm (Rokach and Maimon 2005). Besides the input data (sometimes referred to as feature vectors) the number of clusters

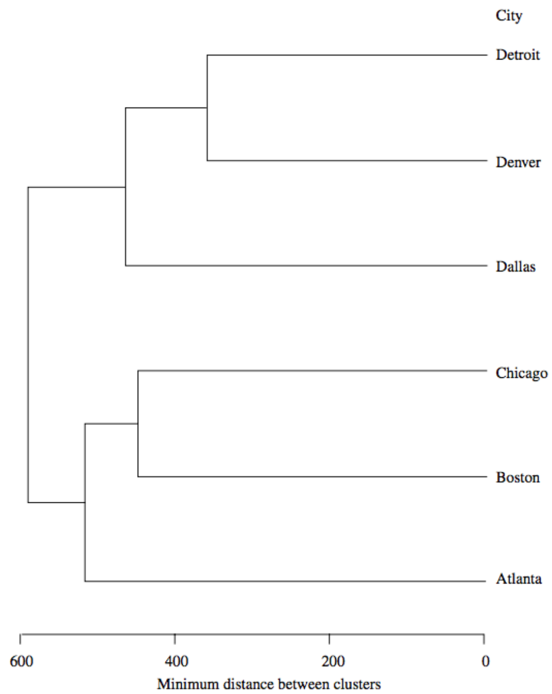


Figure 3.24: Example of a dendrogram as a result of a hierarchical clustering process. It shows the first six observations of a city crime dataset. Single linkage clustering is used. (Rencher and Christensen 2012)

(referred to as *k* in this case) needs to be given. In an initial step, a random solution as vectors of means is created. The input data needs to be assigned to one of the *k* clusters based on a distance measure. This allocation can be used to compute new cluster centres based on the mean of all objects belonging to that cluster (the midpoint). These steps are repeated as long as the result improves (Rokach and Maimon 2005). The k-means algorithm is very fast and thus can be applied to large databases, but has some disadvantages, e.g. the sensitivity towards outliers (Han et al. 2001, Estivill-Castro and Yang 2000). That is why alternative partitioning algorithms such as the expectation maximization algorithm (uses a probability distribution) or the k-medoids methods (uses the medoids instead of the mean) were developed (Han et al. 2001).

Whereas most clustering algorithms are based on the distance between objects, the density based methods use density estimators. Regions of high density are separated from those of low density which is useful in the case of noise filtering (Rencher and Christensen 2012). Additionally, the identification of arbitrary-shaped clusters is possible with this approach (Rokach and Maimon 2005, Han et al. 2001).

The efficiency of density based methods is related to the number of dimensions. To enhance the efficiency, the space can be transferred into a grid structure. This results in a fast processing time which is usually independent of the number of objects. These so called grid based methods only depend on the grid size (the number of cells) in each dimension (Han et al. 2001).

In order to factor in spatial constraints (such as mountains or rivers), constrained based methods were developed (Han et al. 2001). This is only a small excerpt of existing clustering approaches, whereas partitioning and hierarchical clustering are commonly used approaches (Rencher and Christensen 2012). Based on the nature of the input data, the input for a specific algorithm, and the size of the input dataset an appropriate approach can be chosen. Some techniques need numerical values as input whereas others only handle categorical values. Depending on the algorithm, varying information is required in advance, such as the number of clusters for a k-means clustering, or the specification of distance or similarity measures.

A clustering process aims at generalizing the dataset based on similarities. In other terms, the grouping of all objects into one single cluster is the complete generalization of the input data – which is usually not required. Thus, the selection of an optimal number of clusters (the resolution of generalization) is an important task in terms of cluster analysis (Bahrenberg et al. 2003). The aim is to determine the value for g which fits the data best (Rencher and Christensen 2012). A rule of thumb which can be used with hierarchical as well as partitioning methods is the following (Mardia et al. 1979):

$$g = \sqrt{\frac{n}{2}} \tag{3.33}$$

Another common approach is the so called elbow method which uses the percentage of variance explained by a cluster solution. If the additional cluster does not noticeably increase the variance value, the number of clusters should not be increased. This point can easily be identified when plotting the percentage of variance explained against the number of clusters. The percentage of variance explained is defined as the ratio between-group variance to the total variance (Rencher and Christensen 2012). However, this point is can not always be unambiguously identified (Ketchen and Shook 1996). Other possible methods use the information criteria, such as Akaike information criterion (AIC), Bayesian information criterion (BIC), or the Deviance information criterion (DIC), which can compare the results of different clustering outputs in order to determine the optimal value g. In a stepwise clustering approach, such as the hierarchical methods, the distances (or similarities) between a cluster can be useful measures. Mojena (1977) presented a formalization:

$$a_j > \bar{a} + ks_a, J = 1, 2, ..., n \tag{3.34}$$

where:
$a_1, a_2, ..., a_n$ are the distance values for stages with $n, n-1, ..., 1$ clusters,
$\bar{a}$ is the mean of the a's,
s_a is the standard deviation of the a's,
k is a constant.

A k-value between 2.75 and 3.5 is recommended by the author, whereas Milligan and Cooper (1985) suggested a value of 1.25. Another possibility is the use of a partitioning around medoids clustering, which estimates the number of clusters by optimum average silhouette width. The clustering technique is similar to the k-means clustering (it also is a partitional algorithm) but in contrast to k-means medoids instead of centroids are used. It determines the optimal cluster number based on silhouettes (Kaufman and Rousseeuw 1990).

Several other approaches exist (e.g. Calinsky criterion or Affinity propagation) which are not explained in greater detail. Bahrenberg et al. (2003) recommends to use more than one approach in order to find the best compromise between information loss and generalization.

A non-hierarchical clustering approach (k-means) is used in this thesis. The k-means algorithm is used because it is efficient in terms of computing time (Rokach and Maimon 2005). The aim is to develop an approach for determining settlement functions which is applicable to various kind of datasets. This implies data on a small scale as well as huge areas. Additionally the input variables should be interchangeable and of varying quantity, thus it needs to be a data driven workflow with only very limited user interaction. Whereas the latter condition is not hard to achieve (the clustering does not have any limitations in the number of input variables), the former can be a bigger challenge. In a first attempt the measured distance values are used as input for the cluster analysis. Regardless of the clustering method (hierarchical or partitioning) some excavation site are separated into different clusters although the same settlement type can be assumed. That happens due to environmental variables in greater distance which still influence the clustering result even if they are several hundred kilometres apart. As shown in figure 3.25 the excavation sites close to the whitewater lake are assigned to two different clusters, namely cluster 1 and cluster 6. The same happens with the excavation sites close to the blackwater lakes where either cluster 2 or cluster 6 is chosen. Obviously other, more distant environmental parameters are influencing the result. The number of variables – which equals the number of influencing parameters in this study – can have influence on the processing time (Modenesi et al. 2007). Due to that reason, the initial aim was to reduce the number of variables using a principle

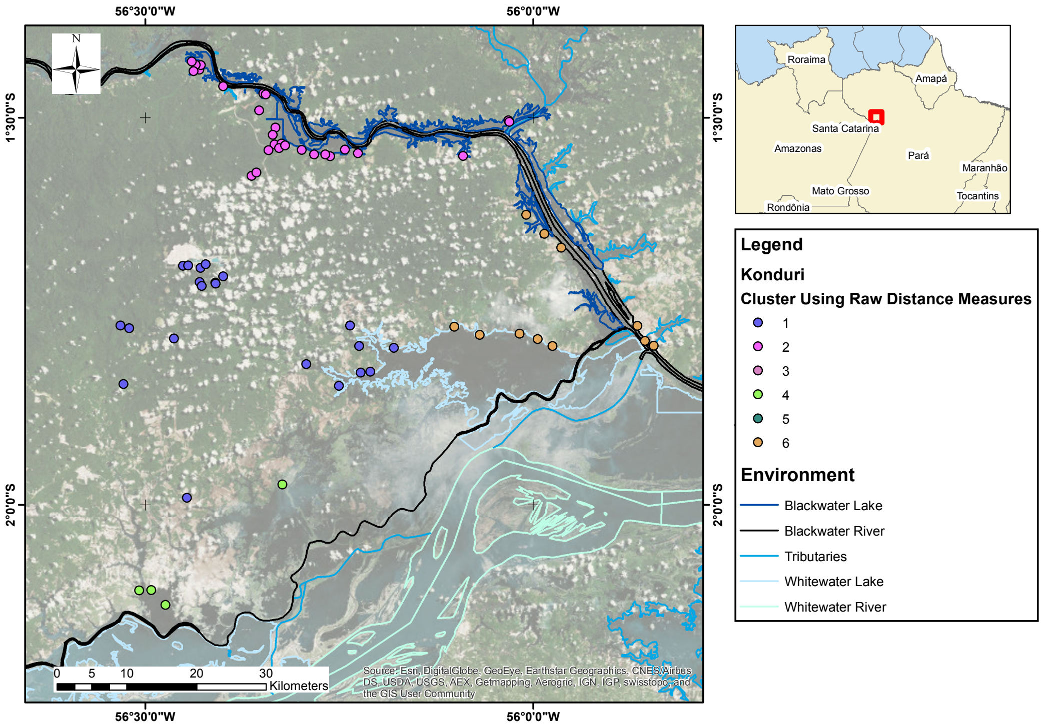


Figure 3.25: Excerpt of the Konduri clustering result if raw distance values are used

component analysis (PCA). PCA is a technique which allows the reduction of variables. It transforms the variables into a smaller set of uncorrelated variables, the so called principle components (Kabacoff 2011, Bahrenberg et al. 2003). The principle components are derived by identifying the line through the data set which has the largest possible variance, and thus accounts for as much variability as possible, whereas the lines have to be orthogonal to the previous components. This makes the PCA an orthogonal linear transformation which derives a new coordinate system in which the data is transformed (see figure 3.26 for a better understanding). Besides the decrease of input variables the PCA also reduces potentially correlated variables to solely uncorrelated variables (Kabacoff (2011). The resulting principle components are used as input for the cluster analysis. The problem is, that the reduction of variables can lead to an unintentional split into two or more clusters. The excavation sites which are far away from any bigger water source (in the western part of figure 3.27) are divided into two clusters. The input data for the PCA were only the distances towards water sources (not all the considered environmental variables), thus a common cluster for all distant locations would be expected. By rescaling the raw distance measures, the influence of distant environmental parameters can be limited. The distance measures were rescaled by applying the following equation

$$dist_{rescaled} = 1/dist * x \tag{3.35}$$

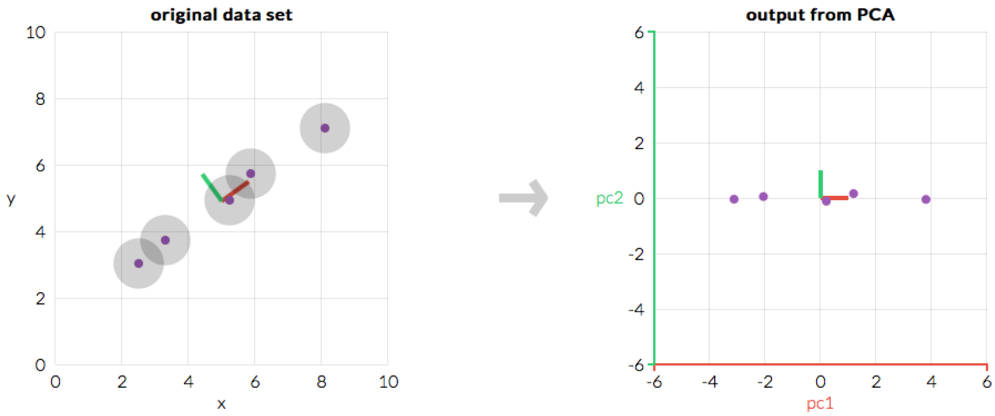


Figure 3.26: Dimensions can be reduced using PCA. As can be seen on the right, there is almost no loss of information if only the first principle components is used in further analysis. (Powell and Lehe 2015)

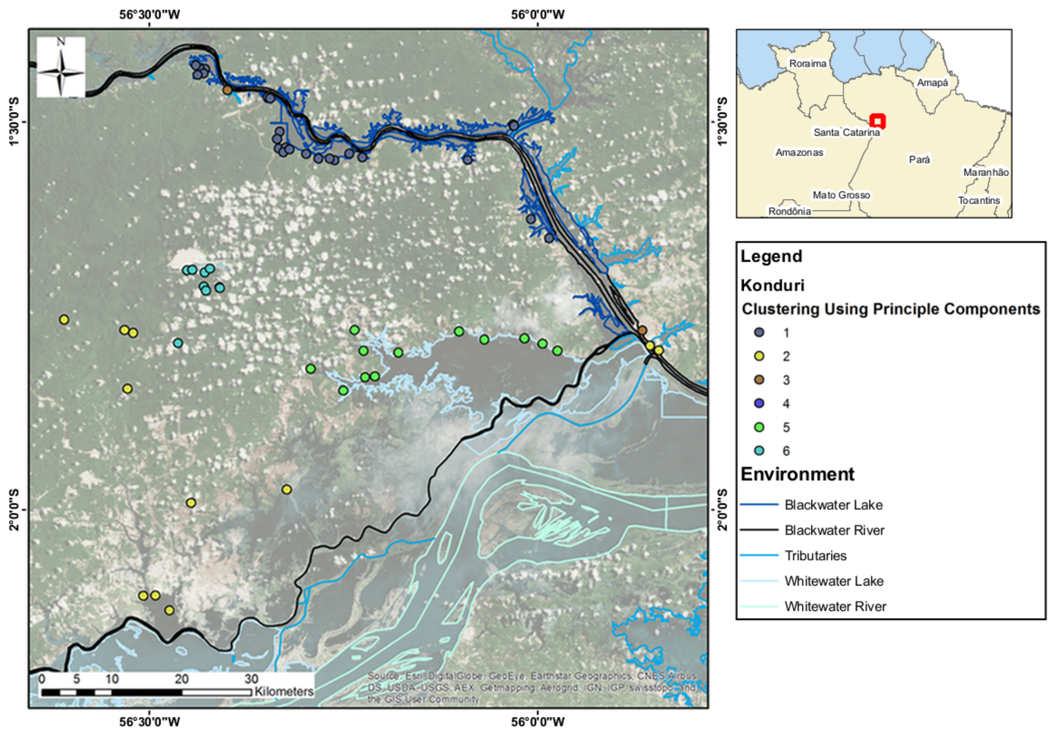


Figure 3.27: Excerpt of the Konduri clustering result if a PCA is applied

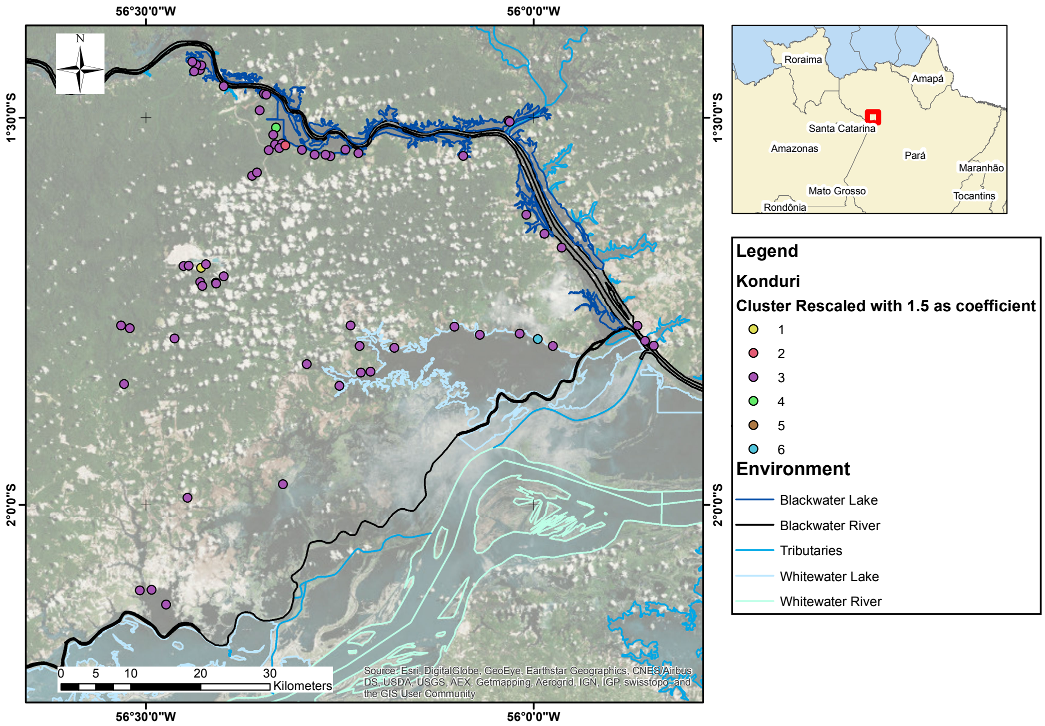


Figure 3.28: Excerpt of the Konduri clustering result if distance measures are rescaled. A value of $x = 1.5$ is used.

where x is a coefficient which can be varied.

The difficulty was to determine an appropriate value for x. By using a small value the same problems occur as with the raw distance values. If a too high value is chosen, the difference of only a few metres led to a different cluster solution. Another disadvantage is that this attempt needs a (possibly subjective) interpretation of the clustering result in order to determine the ̈bestc̈oefficient value x. The figures 3.28 and 3.29 are created using the equation 3.35 whereas figure 3.28 uses s coefficient of $x = 1.5$ and the figure 3.29 uses $x = 0.5$. As can be seen, the result is highly dependant on the coefficient. While a coefficient of 1.5 leads to a cluster result where almost all excavation sites are summarized in one cluster (83 of 88 findings are assigned to cluster 3), the coefficient of 0.5 produces different results. The excavation sites in figure 3.30 are all within a 350 metres distance to the blackwater lake and are seemingly assigned to the same cluster, but are actually assigned to three different clusters. Whereas the excavation site of cluster 7 is within 30 metres of the lake, the excavation sites of cluster 5 are up to 300 metres away. The excavation site which is assigned to cluster 6 is about 350 metres away from the lake. It is probably possible to find an optimal value x but it needs to be determined for each selection of excavation sites. In order to develop a data driven approach

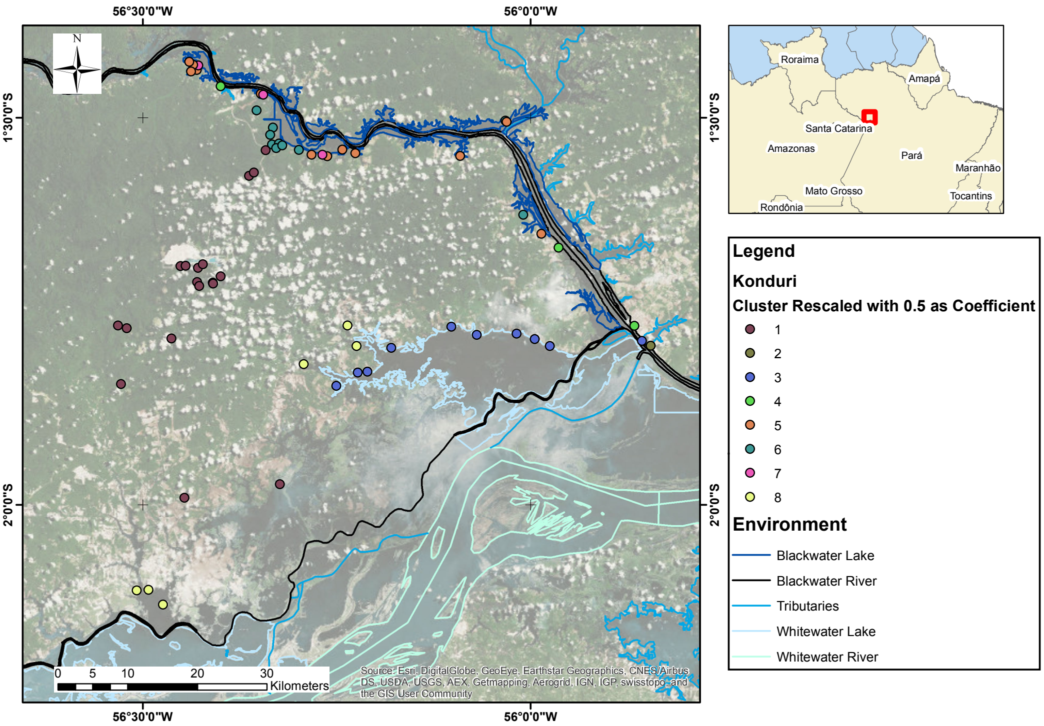


Figure 3.29: Excerpt of the Konduri clustering result if distance measures are rescaled. A value of $x = 0.5$ is used.

neither principle components nor rescaled values can be used.

One premise in this thesis is that the influence of resources have influence on the functional settlement pattern analysis and that the differences in resource availability refer to different settlement functionality. If the resource r_1 is needed at all the settlements but is only available near settlement S_1 it can be assumed, that settlement S_1 supplied that specific resource r_1 and has some kind of trading functionality. Clustering settlement locations and their surrounding (not necessarily environmental) resources can cause problems due to the influence of distant resources. The assumption is that several settlements of the same culture are dispersed along two rivers, whereas the one river is full of fish and the other is not and all other influences are the same across the study area (as can be seen in figure 3.31 on the left). If a common – and it does not matter if a partitioning or hierarchical – clustering method is used, the three agglomerations are separated into three clusters (in this case indicated by the colours green, orange and gray as can be seen in figure 3.31 on the right). This result might be satisfying if different agglomerations are to be identified. In order to derive the different settlement types the surrounding parameters are relevant, thus should be grouped into one cluster. In the example of figure 3.31 the scale is important for the assessment of the result. If the river in the south

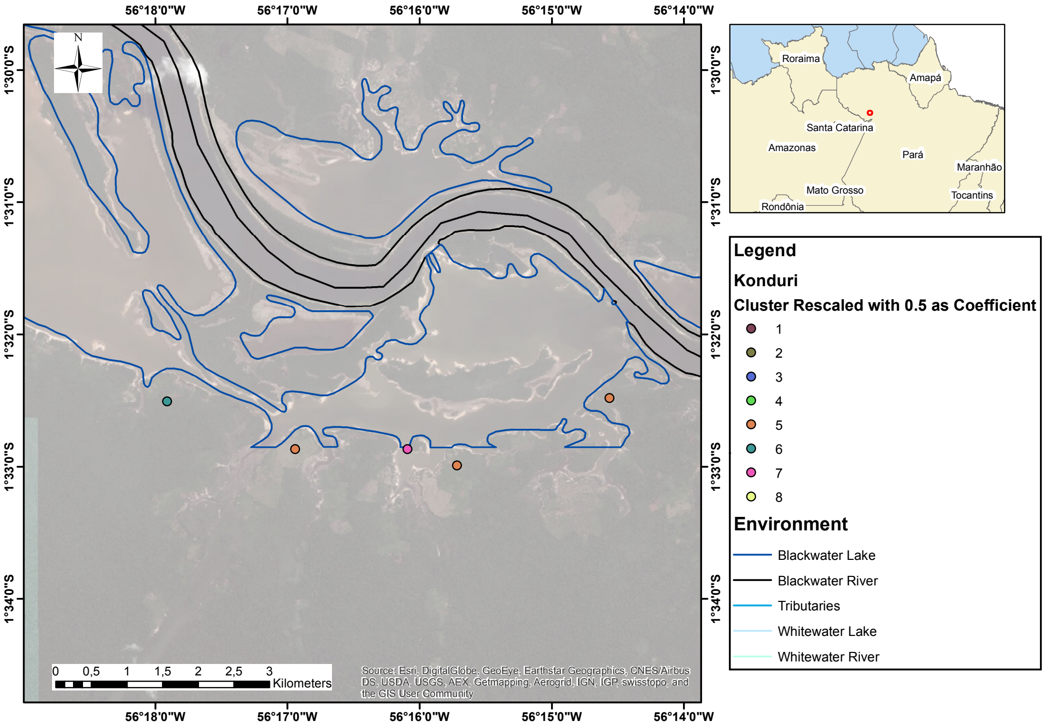


Figure 3.30: Excerpt of the Konduri clustering result if distance measures are rescaled. A value of $x = 0.5$ is used. Problem of sensitivity towards small variances in distance.

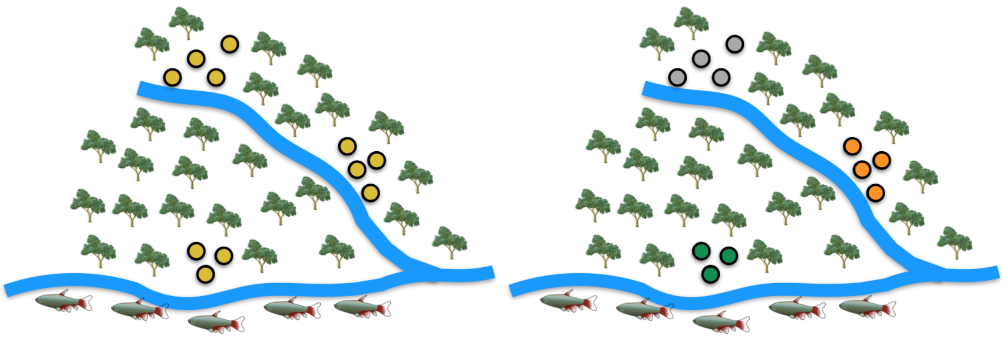


Figure 3.31: Problem of clustering analysis in the case of functional settlement pattern analysis. The left picture indicates possible settlement of one specific culture, the right a possible clustering result.

is within a certain threshold distance and can therefore be seen as food source for all settlements, the river should be considered in the clustering. At a smaller scale this might not be the case. A river which is 100 kilometres away probably had no influence in the settlement decision and thus should not influence the clustering result. It can be assumed that only the resources which were "near" to the settlement were relevant, whereas the meaning of near can vary according to the resource (see also chapter 2.2.2). In order to avoid the unintentional influences of parameters, a maximum distance value max_{near} is used which determines whether a resource is near or not. Thus the aim is to cluster excavation site specific surroundings. This means that everything between the settlement and the maximum distance is considered to be near. But how exactly can nearness be defined and does the perception of nearness vary according to the resource or the abilities of a culture? It is assumed that different needs and capabilities lead to different distance values why this value can be individually set for each parameter. As mentioned above (see e.g. chapter 2.2.2 research about cultures and their willingness to overcome distances exists. Nevertheless, it is a very complicated and sometimes impossible task to determine which max_{near} to use. Due to this the methodology can be applied various times in order to calculate different scenarios. It is further assumed that the distance varies according to the resource as well as the settling culture. Due to the lack of information about the max_{near} values an inner and outer boundary ($max_{near_{outer}}$ and $max_{near_{inner}}$) of max_{near} values is used. This allows a range of max_{near} values rather than just one single value. The methodology also allows to set these values individually for each culture. This allows the assignment of varying $max_{near_{outer}}$ and $max_{near_{inner}}$ values for the same location but different settlement periods (also see table 3.3). These boundary values now determine whether a resource r_i within a specific dis-

settlement_boundary_id_pk	culture_id_fk	max_boundary_inner	max_boundary_outer	parameter_type
1	1	500	5000	wood
2	1	300	1000	water
3	1	2000	10000	stones
4	2	1000	2000	wood
5	2	50	750	water
6	1	500	5000	wood
7	1	500	1500	water

Table 3.3: Table which defines the inner and outer boundary values $max_{near_{outer}}$ and $max_{near_{inner}}$ for the clustering analysis. The field *culture_id_fk* determines the culture for which the values are set.

tance should be considered in the cluster analysis for this settlement. One example is given in figure 3.32 where $max_{near_{outer}}$ and $max_{near_{inner}}$ are symbolized with two different buffer sizes. If $max_{near_{inner}}$ is chosen as maximum distance, only the river close to the settlement is considered in the cluster analysis. In contrast to that, the second river (which is rich in fish) is factored in if $max_{near_{outer}}$ is used. Based on the used values the results of the cluster analysis vary. A boolean value serves as input for the cluster algorithm, the measured distance is either near (within 0 and max_{near}) or not ($> max_{near}$). In order to derive the most likely clustering solution $max_{near_{outer}}$ and $max_{near_{inner}}$ are used as approximate values. A high number of clustering analyses (e.g. 1000 times) is performed using a random value between $max_{near_{outer}}$ and $max_{near_{inner}}$ for each parameter (and settlement and culture if specified). Several cluster comparison methods exists, but can vary according to the focus of the comparison. Asymmetric measures are used if clustering results should be compared in relation to an optimal clustering solution (e.g. the Meila-Heckermann-Measure (Meilă and Heckerman 2001)). The Fowlkes-Malows Index (Fowlkes and Mallows 1983) or the adjusted Rand Index (Hubert and Arabie 1985, Rand 1971) compare a cluster solution to its expected value under the null hypothesis. Strong assumptions are made for the null hypothesis – such as the fixed number of cluster sizes – which can be violated if compared with the result of a clustering algorithm (Wagner and Wagner 2007). In this case, the consensus clustering approach is used, which does not return one similarity value but rather information about quality measure for each item and cluster. Consensus clustering was initially invented by Monti et al. (2003) to unsupervised estimate the number of classes. Regardless of the used clustering algorithm consensus clustering allows the comparison of several clustering runs (Wilkerson and Hayes 2010). It is based on a consensus matrix m_{cons} which stores,

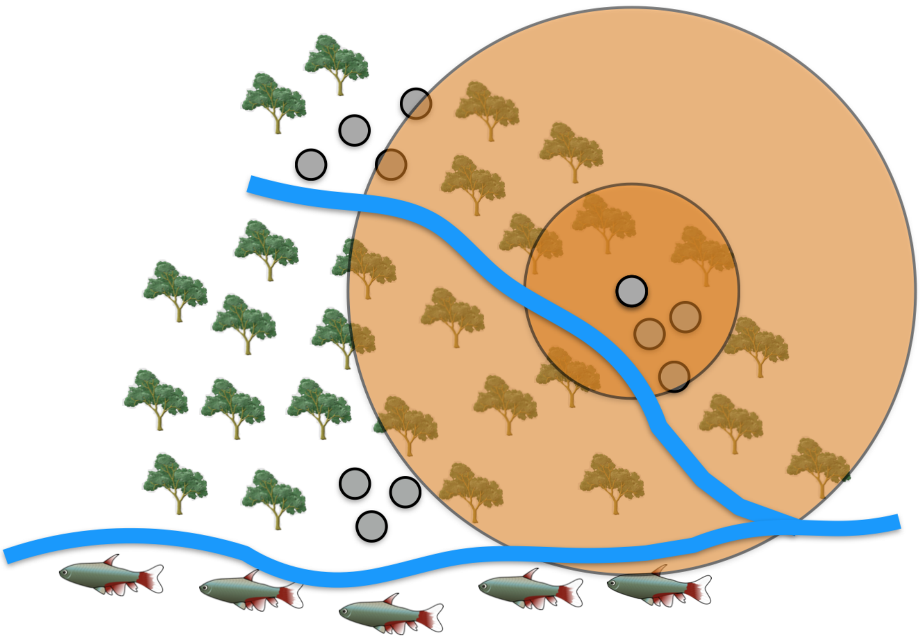


Figure 3.32: Based on the inner and outer boundaries $max_{near_{outer}}$ and $max_{near_{inner}}$, the river in the south is factored in or not

> for each pair of items, the proportion of clustering runs in which two items are clustered together. (Monti et al. 2003)

The matrix m_{cons} is a $N * N$ matrix, whereas N represents the number of items in a dataset and is based on the average of all connectivity matrices m_{conn}. Each connectivity matrices m_{conn} represents the result of one clustering run and is defined as follows:

$$m^h_{conn}(i,j) = \begin{cases} 1 \text{ if items i and j belong to the same cluster,} \\ 0 \text{ otherwise.} \end{cases} \tag{3.36}$$

where:
h h-th cluster run,
i and j two items (settlements in this case).

An additional indicator matrix m_{indi} – which is also a $N * N$ matrix – is used to properly normalize the values of the connectivity matrices m_{conn}. It is defined as follows:

$$m^h_{indi}(i,j) = \begin{cases} 1 \text{ if both items i and j are present in the Dataset } D^h, \\ 0 \text{ otherwise.} \end{cases} \tag{3.37}$$

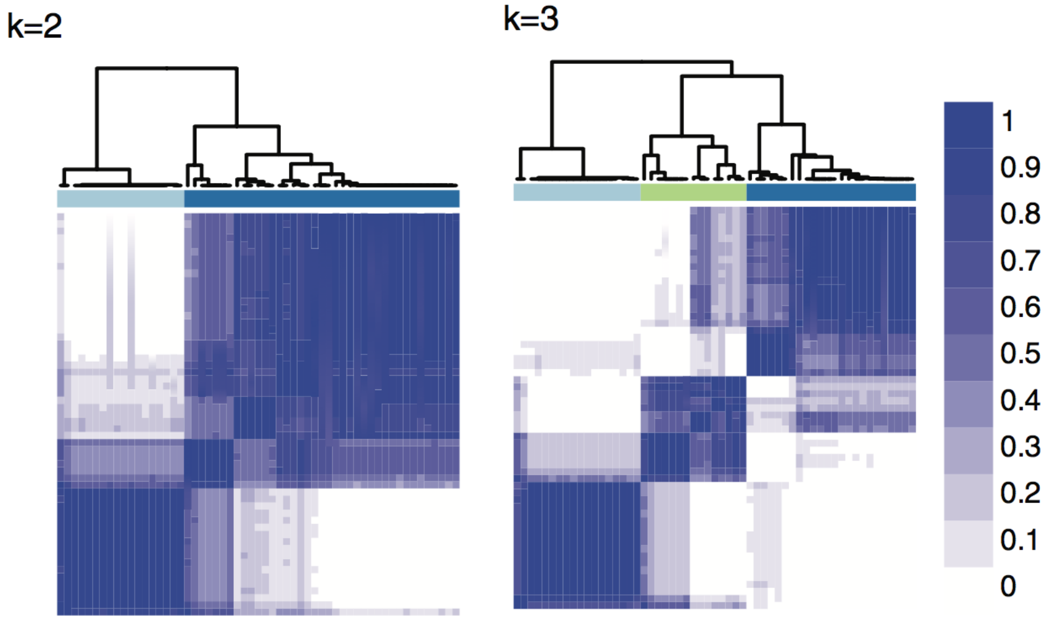


Figure 3.33: Visualization of m_{cons} combining a dendrogram and color coded values of m_{cons} with two (on the left) respectively three (on the right) clusters. A value closer to one (dark blue) indicates that this entry (i, j) is often grouped in the same cluster. The colors below the dendogram show the cluster membership. (Wilkerson and Hayes 2010)

whereas D^h is the used dataset in the h-th run. The indicator matrix m_{indi} is important if not all items of one dataset are included in all cluster runs. Based on the matrices m_{conn} and m_{indi} the consensus matrix m_{cons} can be calculated.

$$m_{cons} = \frac{\sum\limits_{h} m^h_{conn}(i, j)}{\sum\limits_{h} m^h_{indi}(i, j)} \tag{3.38}$$

That means that the entry (i, j) in the consensus matrix m_{cons} counts how often i and j are clustered together and divides the value by the total number of times both are selected (Monti et al. 2003). The term consensus index refers to the entry (i, j) in m_{cons}. To facilitate the analysis of the consensus clustering process, the entries in m_{cons} can be arranged so that the items of one cluster are adjacent to each other (see also figure 3.33). More generally spoken, the result is good if non-overlapping blocks of value one (always together in the same cluster) are surrounded by zeros (never together in one cluster). In the case of a singleton clustering (a clustering where the number of clusters equals the number of items) would – based on the definition above – also be considered to be a good clustering solution, which is usually not the desired result since one aim of clustering is to reduce the number of variables. Two additional values help to assess the clustering result, namely the item

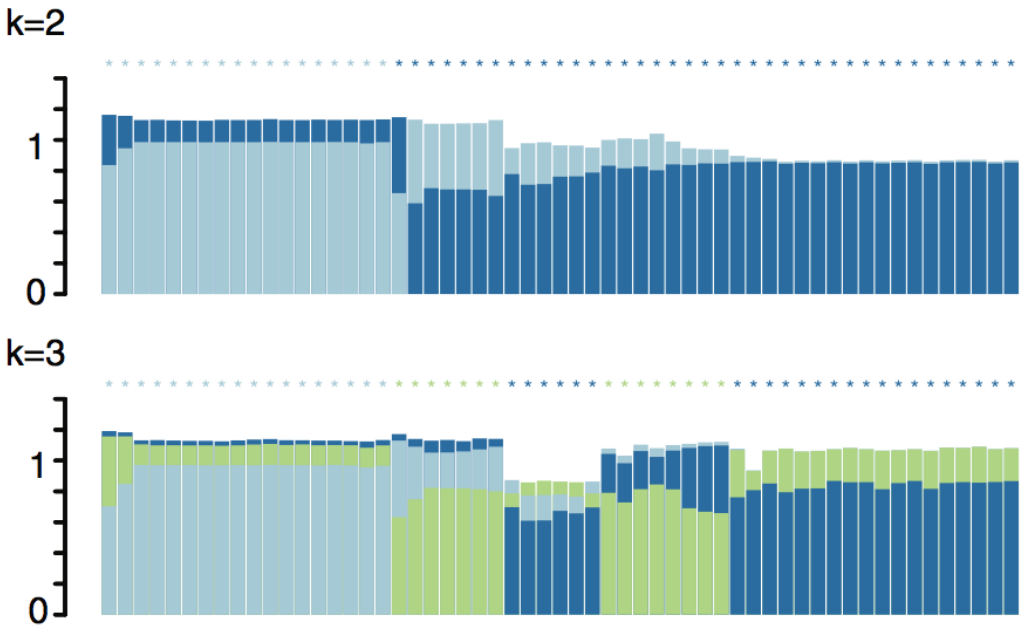


Figure 3.34: Visualization of the item consensus values for two (on the top) respectively three (on the bottom) clusters. The coloured asterisks indicate the associated consensus cluster. (Wilkerson and Hayes 2010)

consensus value and the cluster consensus value. The former is the average value (also referred to as consensus value) between an item and members of a cluster (also referred to as consensus cluster). This allows to determine the cluster with the highest value (and thus are highly representative for a cluster), as well as identifying items which have mixed cluster association (see also figure 3.34). In contrast, the cluster consensus value is defined by the average pairwise item consensus of items in a consensus cluster. This allows to assess the impact of a new clusters on the cluster consensus values of existing clusters (Wilkerson and Hayes 2010). Due to these values consensus clustering allows to validate the clustering results and gain confidence about the significance of them. The consensus clustering is adapted in order to compare the results of the many cluster analyses of the settlements. Thus m_{conn} counts the pairwise clustered items for all cluster runs (and not different cluster numbers) and is normalized using m_{indi}. This allows to identify the clustering solution with the highest consensus for all settlements. This clustering solution with the highest consensus value can then be used in order to determine approximate settlement type related nearness values.

3.2.4 Territorial Analysis

Site catchment as well as territorial analysis are the two archaeological concepts in order to identify former areas of influence (see chapter 2.2.2). Site catchment

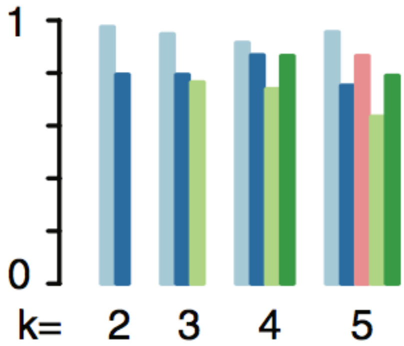


Figure 3.35: Comparison of different cluster runs using the cluster consensus value (Wilkerson and Hayes 2010)

analysis is an empirical approach which is based on the findings whereas territorial analysis is a theory driven approach (Renfrew and Bahn 2005). Site catchment analysis needs more detailed information about the excavation sites and the availability of resources, which are not existent for each data set. One major goal of this research is to maximize the number of usable datasets (by limiting the needed information) and to minimize expert interaction (by using statistical methods rather than expert knowledge). It is for that reason that a territorial analysis rather than a site catchment analysis is performed.

> Common GIS packages will offer at least two basic tools useful in territorial modeling: geometric partitioning using Voronoi diagrams and the calculation of cost surfaces. These are the basic tools for creating territorial GIS models from the ground up. (Ducke and Kroefges 2007, p. 247)

In the case of functional settlement pattern analysis modelling geometric partitioning methods have some disadvantages, especially when working with incomplete datasets (see also chapter 2.2.2). Therefore a cost surface modelling approach is used which better matches the needs. Various approaches for creating a cost surface exist. A commonly used approach is to simply use the slope and derive the cost attribute based on time (e.g. based on the walking speed which is influenced by the slope) (e.g. de Smith 2006, De Silva and Pizziolo 2001, Van Leusen 1999). A more advanced approach is to implement a transportation network in order to derive the accessibility (e.g. Ueberschär 2013, Delamater et al. 2012, Huerta Munoz and Källestâl 2012, Schuurman et al. 2006, Tanser et al. 2006). As mentioned before, it is assumed that some resources were more important than others and people were more likely to overcome greater distances for these specific resources than for others. That means that the cost surface methods given above are not sufficient and other – more complicated – techniques need to be used. In this case, the cost surface that should be based on the influence of such resources in combination with a distance measure rather than only on walking distances (or other cost attributes). Multi-criteria evaluation (MCE) (sometimes also referred to as multi-criteria analysis) techniques can be used in order to derive such a suitability surface (Carver 1991). The idea is to combine multiple criteria in order to identify a suitable solution (Janssen and Rietveld 1990). In terms of spatial analysis this approach allows to classify suitable areas based on the input parameters (criteria). The different input parameters are used as (weighted) overlays and are merged (Bonham-Carter 1994). In terms of data driven (and more inductive) analysis MCE has a drawback. Depending on the chosen merging method, it is required to assign weights or fuzzy membership values to the input parameters. This is usually done by experts which would be opposed to to the inductive approach. In this case another way must be used.

Besides the importance of environmental parameters the variation of these influences based on the settlement function should be considered. Therefore the item consensus values in combination with the related distances to the variables is used. The distances of the relevant variables of all excavation sites are measured and assigned to a fuzzy membership function based on the item consensus value (as shown in figure 3.36). A fuzzy membership function indicates the strength of a membership to a set with values between 1 and 0. The strength of membership decreases with a decreasing membership value. This means, the higher the item consensus value, the the stronger the membership of the specific environmental variable for that specific cluster and distance. It is assumed that the distance values increase or remain the same with an decreasing item consensus value. Therefore multiple buffers are assigned around each environmental parameter whereas the distance is derived based on the fuzzy membership function. Additionally, the item consensus value is assigned to the buffer. Based on this buffers, a suitability surface is created for each cluster resulting in a collection of settlement function related suitability surfaces. Those are built using the variable importance values of the MEM in combination with all the settlement type related fuzzy membership functions. This is done using map algebra (Tomlin 1991).

The suitability surfaces S_{clust} can then be used in order to calculate the territory of the observed culture. In order to calculate this area, a maximum distance $Dist_{max}$ needs to be set, which determines the maximum cost CD_{max}. The maximum distance a culture is willing to overcome in order to exploit the resources vary depending on the analysed culture (as described in chapter 2.2.2). The $Dist_{max}$ value is de-

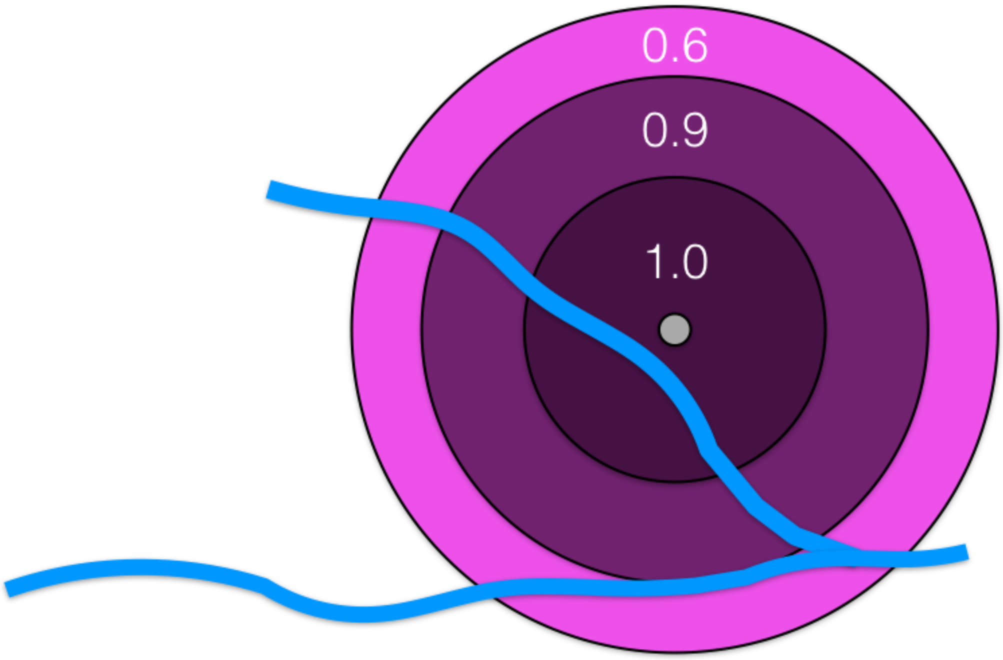


Figure 3.36: Applying item consensus values to the excavation sites based on the settlement function and distance measures. The values indicate the item consensus values whereas the different buffer sizes are the measured distances regarding the item consensus value.

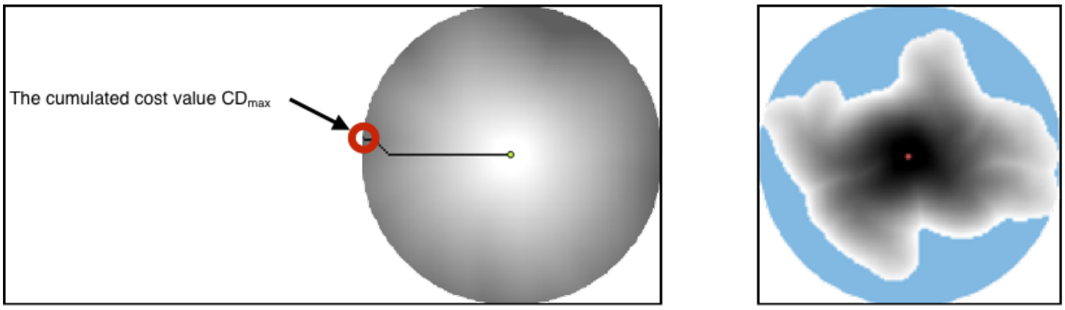


Figure 3.37: Left: Calculation of the cost distance value CD_{max}. Right: Calculation of the territory based on the maximum cost of CD_{max}. (adapted from von Groote-Bidlingmaier et al. 2014a,b)

termined by the maximum distance towards an influencing environmental variable. In order to determine the maximum cost value CD_{max} and calculate the territory, the following steps have to be performed:

1. Buffering each excavation site with its maximum distance value $Dist_{max}$.
2. Calculating the least cost path between the excavation site and the outline of the buffer.
3. Get the maximum least cost path value of all excavation sites for the specific cluster. The resulting maximum accumulated least cost path value equals the value CD_{max} (as shown in figure 3.37).
4. Creating the territories based on the cost surface S_{clust}. Therefore the cumulative cost value CD_{max} is used as maximum cost value.
5. Merging the results to represent the (sometimes connected) territories of a culture.

4 Case Study – Gaining Knowledge Based on the Environment

4.1 The Amazon Basin and its Environment

The largest socio-geographic division in Brazil is the Amazônia Legal which contains nine states. Besides the seven states of the North Region (namely Acre, Amapá, Amazonas, Pará, Rondônia, Roraima and Tocantins) it also covers parts of Mato Grosso and Maranhão. The division contains all states of the Brazilian part of the Amazon basin. Its main characteristic are the tropical climate (hot and humid) and vegetation types, i.e. the tropical rain forest. The Amazon basin is bordered by the Brazilian Highlands (in the south), the Guiana Highlands (in the north) and the Andes (in the west) (Sioli 1983). Tertiary sediments of the Amazon sink cover the area in between these two shields. The geological history leads to an unequal distribution of stones.

Due to the region's heavy rainfall some areas are regularly flooded in the rainy season. The Amazon river has shallow banks and its riverbed is located within an alluvial plain (várzea) which is usually completely flooded in the rainy season (Meggers 1984). Three different river types can be distinguished in the Amazon basin, namely the whitewater, clearwater and blackwater river types. Due to differences in the chemical composition, based on the different headwaters, these types differ in the availability of food (because of the different availability of nutrients). The whitewater rivers, basically the Amazon river and some tributaries, are rich in nutrients whereas the clear- and blackwater rivers are nutrient-poor. Although differences between clear- and blackwater rivers exists, those two types are often treated equally (e.g. Heckenberger and Neves (2009) only distinguishes between white and clearwater river systems). A similar distinction is used within the scope of this research.

The Amazon region has two different landscape types, várzea and terra firme. Meggers (1984) stated, that the terra firme consists of all land between the rivers and covers about 98% of the Amazon. Most of the terra firme is covered with tropical rain forest and is not exposed to the changes in water level (Sioli 1983).

4.2 Archaeological Model – Expected Settlement Patterns of the Observed Cultures

> Archaeological research, as well as every scientific discipline, is based on the dual intellectual process of observation and description. It is through observation and integration into models that human knowledge is developed.(Deravignone and Macchi Jánica 2006a, p. 121)

As mentioned in chapter 2.2 the environmental parameters used in archaeological settlement pattern studies vary. The focus usually is on parameters which are related to food availability and production capacity. Table I tries to categorize these parameters based on previous archaeological settlement pattern studies. Some publications do not specify the variables in detail which made classification difficult (e.g. climate can be measured in various ways and cannot be described in one variable). Based on the commonly used variables and in agreement with Klaus Hilbert the following parameters are chosen:

- white water rivers,
- black water rivers,
- tributaries,
- white water lakes,
- black water lakes,
- waterfalls,
- difference in altitude to closest river,
- slope,
- precipitation,
- vegetation type,
- relief type,
- soil condition and
- soil type.

Whereas the last items are quite common in other analysis, the various water types were used in order to see if there is a difference in the functional settlement pattern according to the water type. The assumed importance of rivers in the Amazon is underlined by two drawings published by Koch-Grünberg (1923) which depicts rivers and its tributaries drawn by two Indians. Both maps show a detailed overview of the river networks and indicate that the knowledge about the network is important.

As mentioned above, white water is rich in food and thus is assumed to be more attractive. Lakes usually have a higher availability of food. Additionally they provide protection because they are usually not visible from the main river network. Another reason is the importance of rivers for transportation and connectivity. Even though some settlements were connected by paths which were used for generations (Nimuendaju 1952, Koch-Grünberg 1923) the river network was used as a transportation route (e.g. Denevan 1996, Koch-Grünberg 1923). Settlements usually were located next to the main channels or at least close to rivers which were connected to the main channel (Denevan 1996). In order to consider these differences in settlement location selection, the distance to tributaries is a separate variable. The waterfalls are important due to the availability of stones which are not ubiquitous in the Amazon. Long journeys were necessary to get to the waterfalls (they are located at the bounding Guiana and Brazilian highlands) (Hilbert 1977). Various authors assumed (e.g. Denevan 1996, Koch-Grünberg 1923) that the excavation sites are located on the bluff zones adjacent to the river. Thus, the altitude between the excavation site and its closest river is important.

Most of the data is provided by the Instituto Brasileiro de Geografia e Estatística (IBGE). The waterfalls were located manually using aerial images because IBGE does not provide the data. The slope is derived using a digital elevation model (DEM) , which is based on SRTM (Shuttle Radar Topography Mission) data. The river network is provided by IBGE but the classification into the different categories needed to be done manually due to the absence of river type information. The relevant water type can be seen on aerial photographs and can even sometimes be identified based on the river name (e.g. Rio Negro). The classifications of soil type, soil condition, temperature, precipitation, relief and vegetation are inherent in the data – provided by IBGE – and are not changed (see figures 4.1, 4.2 and 4.3 for the classification). Changes in the input data were made only in case of reservoirs and dam constructions. The units *Average* and *Indiscriminate* in figure 4.3 were originally referred to as *média* and *indiscriminado*. *Média* is used in case of a mainly mixed soil condition with no predominant type. *Indiscriminado* is used for not precisely determined soil conditions. The original soil taxonomy is maintained but a

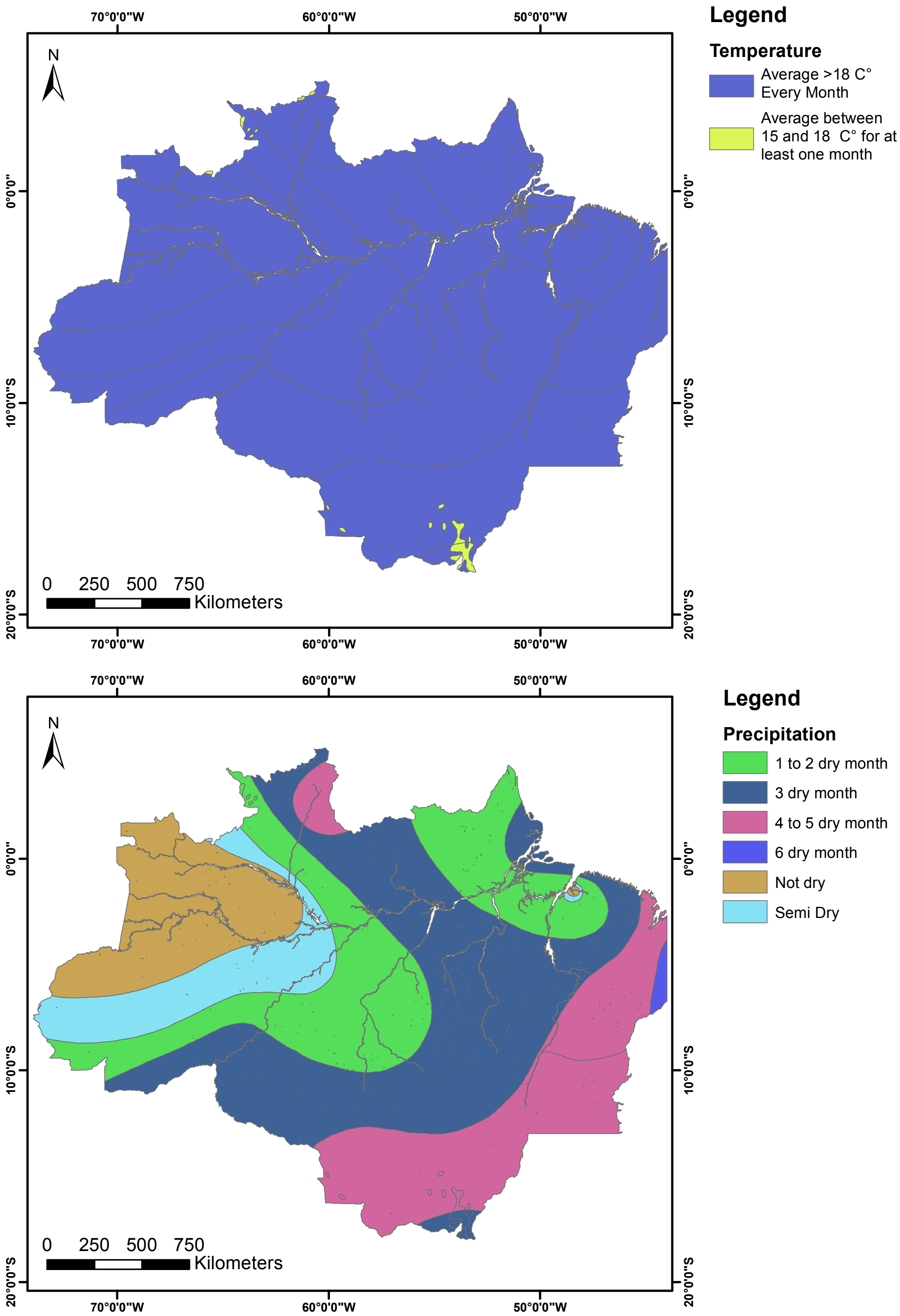


Figure 4.1: Overview of used climate data in the Amazon (source IBGE (2016))

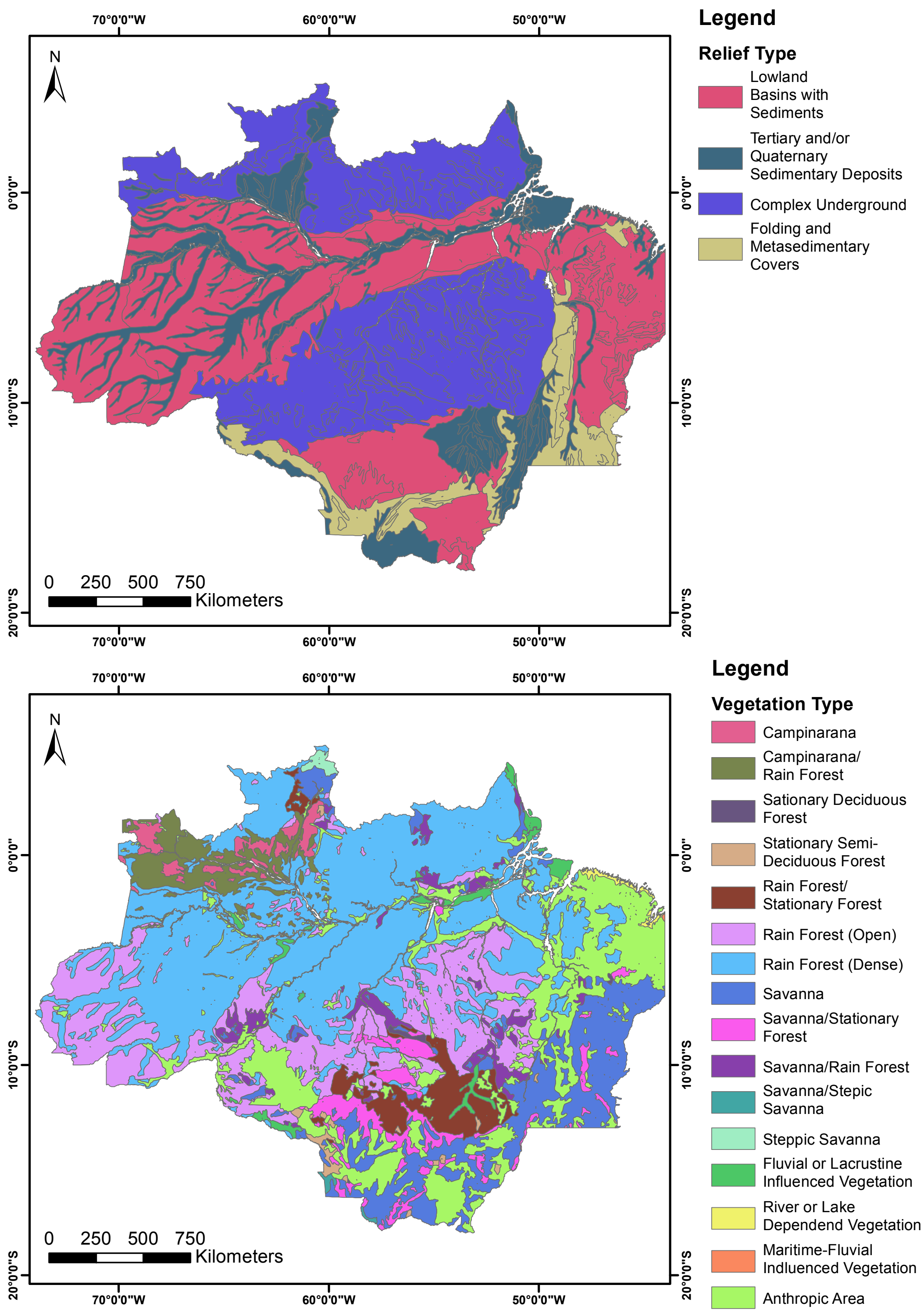


Figure 4.2: Overview of used relief and vegetation data in the Amazon(source IBGE (2016))

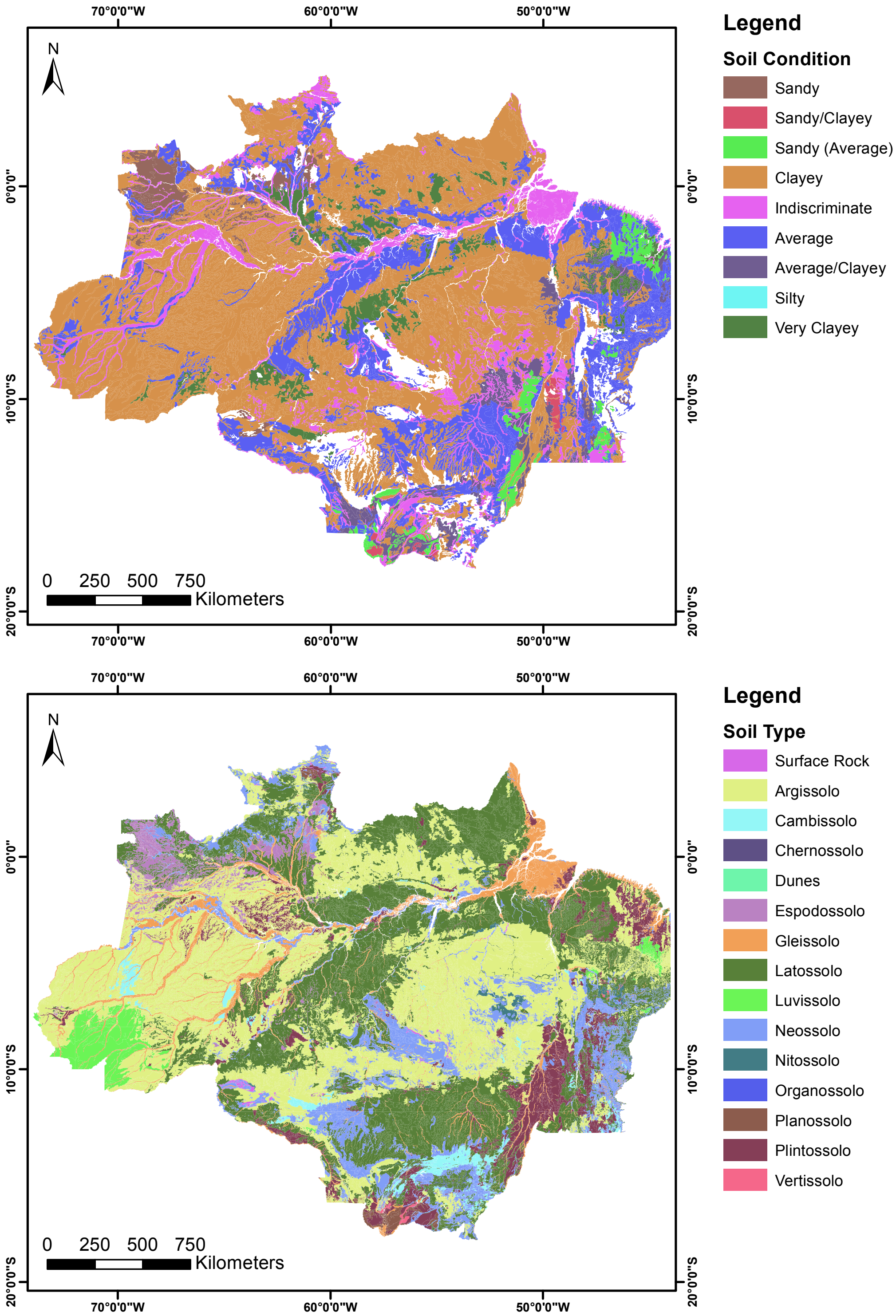


Figure 4.3: Overview of used soil data in the Amazon (source IBGE (2016))

reference to other taxonomies is possible (a more detailed description is given in da Costa and Nanni 2006). The unit *Complex Underground* is initially referred to as *embasamentos em estilos complexos*. It can be seen as bedrock in manifold, complex forms. The unit *Campinarana* is not translated due to a missing counterpart. It can be described as an open park like landscape with a small number of trees.

As can be seen in figure 4.1, the temperature is classified as hot with an average temperature above 18 C°year-round for all excavation sites. It can safely be assumed that the temperature was not influencing the functional settlement pattern, therefore this dataset can be ignored in further analysis.

As mentioned in chapter 2.2 the environmental variables do not necessarily need to be located at the settlement itself but can be a few metres/kilometres apart. It is important to know the resources at the settlement location as well as close to it. Due to that, the distances to each environmental variable (except altitude, precipitation and slope) are measured and assigned to the excavation site.

It can be assumed that different subsistence strategies led to differences in the exploitation of resources and thus afford varying environmental parameters. Due to that, the case study uses the Konduri culture as well as the Guarita culture as an example. One reason is a relatively high number of findings (88 findings which are assigned to the Konduri culture and 69 finding which are assigned to the Guarita culture). Another reason is the period of occurrence as well as the assigned tradition. Both cultures lived roughly at the same time from about 900 BP until the first contact with the Europeans (Heckenberger and Neves 2009, Heckenberger et al. 1999), whereas Konduri is part of the Incised-Punctate tradition and Guarita is part of the Polychrome tradition (Heckenberger 2008, Heckenberger et al. 1999, Hilbert 1968). Besides the differences in ceramic styles the spatial distribution is different. Konduri sites are typically located at the Rio Trombetas and Rio Jamundá whereas Guarita sites are located more to the west at the Rio Negro and Rio Madeira and seem to cover a greater area.

4.3 Technical Setup

As mentioned in chapter 3.1.2, a database is used to store the data. In this case a PostgreSQL database with its spatial extension PostGIS is used. The data is kept on a server and thus avoids the use of local storage. The use of a database server has the following advantages. The data can be accessed from every computer which has the permission to access the server. This also means that the data can easily be provided on a website. The usage of proprietary data formats (such as the file and personal geodatabase from esri) obliges the usage of a specific GIS (ArcGIS in the case of the mentioned formats), whereas the usage of spatially enabled databases allows the use of any GIS, if a connection to a database is supported. In the case of PostGIS many geoprocessing functions are embedded in the database and theoretically no additional software is needed. There are two reasons why an additional GIS software is used in combination with the database. The first reason is that no spatial data viewer is embedded within the database – only the attributes can be seen. By establishing a database connection in a GIS the data can be displayed. The other reason is that some of the raster functionality is not provided in the used PostGIS Version (PostgreSQL version 9.3.5 with the PostGIS extension version 2.1.4). Due to that, Quantum GIS (version Ẅien2̈.8) is used as a viewer and Grass GIS is used (version 6.4) for raster processing. The statistical analysis is done using the statistic software R (version 2.15.1).

The whole process (which comprises all database interactions, geoprocessing as well as the analysis of geodata) is controlled using python (see figure 4.4 for a schematic representation). This requires a couple of python packages which extend the functionality of the python library. The Geospatial Data Abstraction Library (GDAL) package (version 1.11) is a python package for geospatial raster and vector data manipulation. In order to read data from an (Excel-) file the xlrd-package (version 0.9.2) is used. The communication between python and the PostgreSQL database (including the spatial extension PostGIS) is done with the python package psycopg2 (version 2.4.5). The communication between python and R is done using rpy2 (version 2.4.4). Grass GIS as well as Quantum GIS have an embedded python library, thus those libraries are available after the installation of the software.

4.4 Finding settlement patterns – A Statistical Analysis

One assumption in detecting settlement patterns is that various settlements of one culture which occur simultaneously fulfil different functions (as was already mentioned in chapter 3.2.3). This may be due to the availability of resources, trading relations, defence strategies, rituals, etc.

> It is anticipated that one of the most important cultural variables that can be used is the logistic position of the archaeological site itself. It has been shown by many researchers that the position of a settlement in a logistic network determines to a large degree its size and duration of occupation (Verhagen et al. 2007, p. 207)

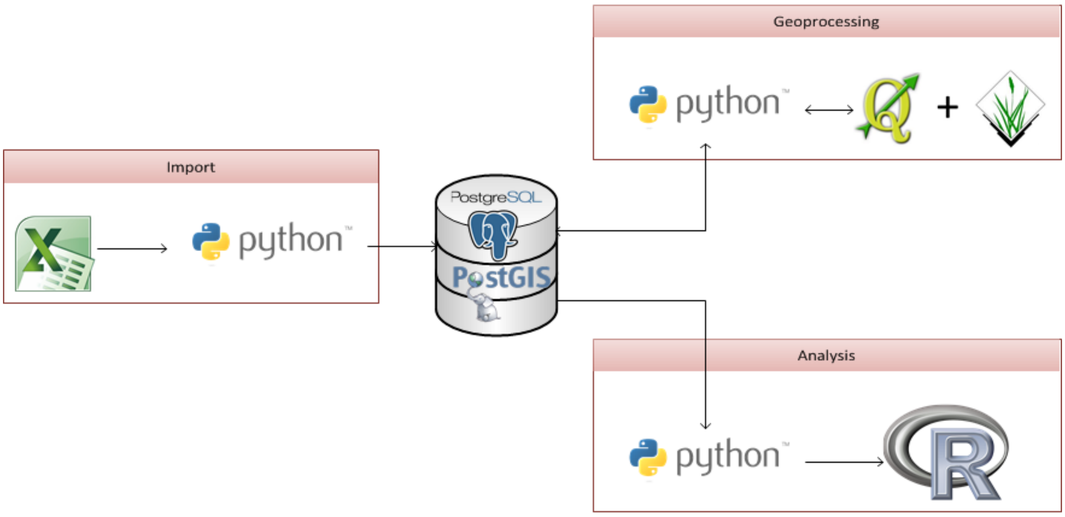


Figure 4.4: The geoprocessing workflow. The excavation sites are stored in a Excel file and imported into a PostGIS enabled PostgreSQL database. The geoprocessing as well as the database handling is controlled using python scripts.

Thus the environmental variables and its variations near the locations are further analysed.

The figure 4.7 shows a boxplot of all observed excavation sites. Each environmental variable is shown on the x-axis whereas the distance towards these environmental variables is plotted on the y-axis. The cultures Konduri and Guarita are plotted in two different colors. The figure 4.8 is similar to the figure 4.7 but only the environmental variables within a 50 kilometres distance are depicted. The figure 4.9 differs from the others. This figure shows two plots, namely a histogram and a boxplot of the slope distribution as well as the altitude distribution.

Figure 4.5 shows the locations and distribution of the known Konduri settlements. An overview of all variables and their distances to the excavation sites is given in figure 4.7. Many of the considered variables occur only in distant areas. When looking at the minimum distances a gap can be identified. Variables are either within a 50 kilometre distance (at least a few excavation sites) or are at least 220 kilometres away. Based on this 18 of the initial 42 environmental distance measures are excluded from further analysis (see figure 4.8 for the boxplot of the selected variables). The overview of the settlement distribution of the Guarita culture is given in figure 4.6. The distances towards the environmental parameters is shown in figure 4.7 and 4.8. As with the Konduri culture some variables seem to be within a certain distance whereas the other environmental parameters are further away. The environmental parameters within a distance of 50 kilometres are very similar to the

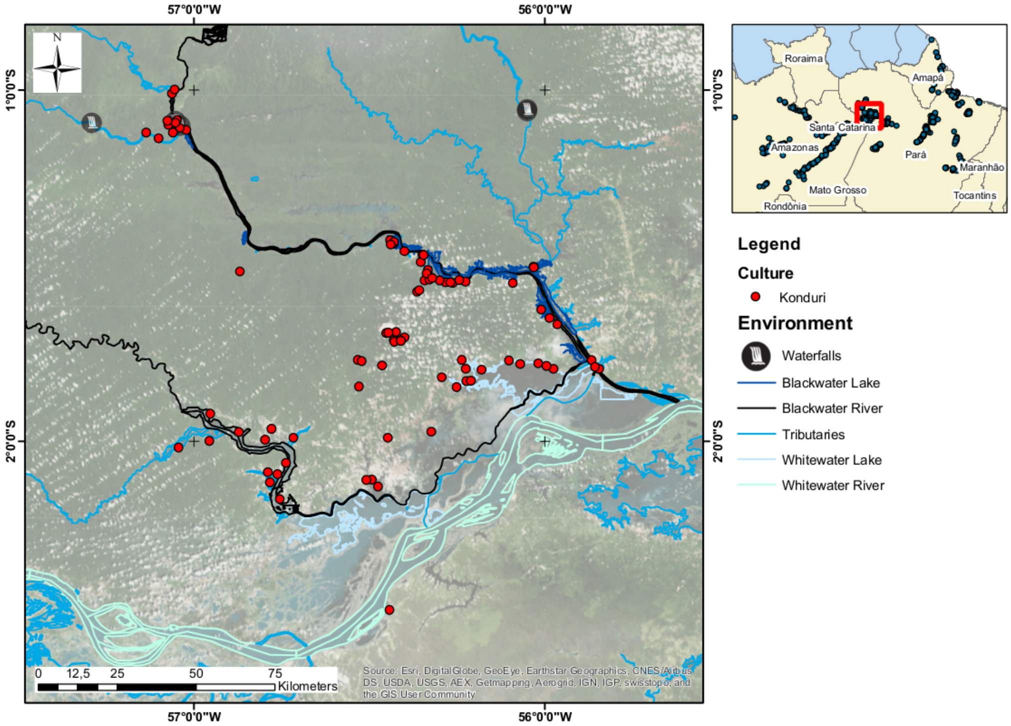


Figure 4.5: Overview about the locations and distribution of the Konduri culture

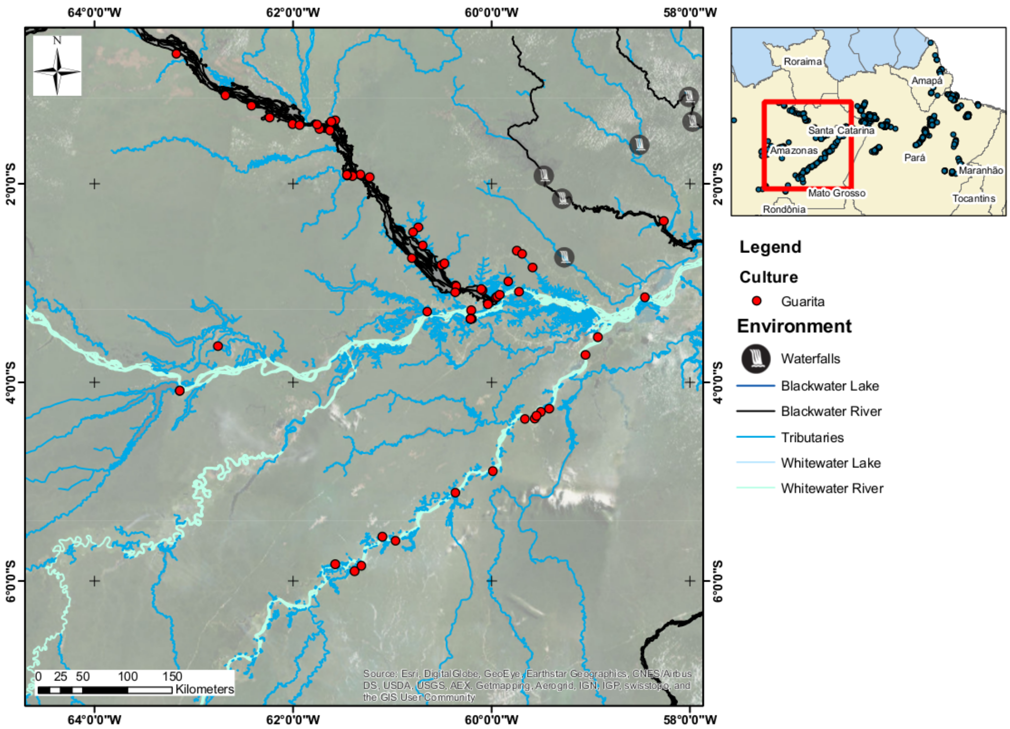


Figure 4.6: Overview about the locations and distribution of the Guarita culture

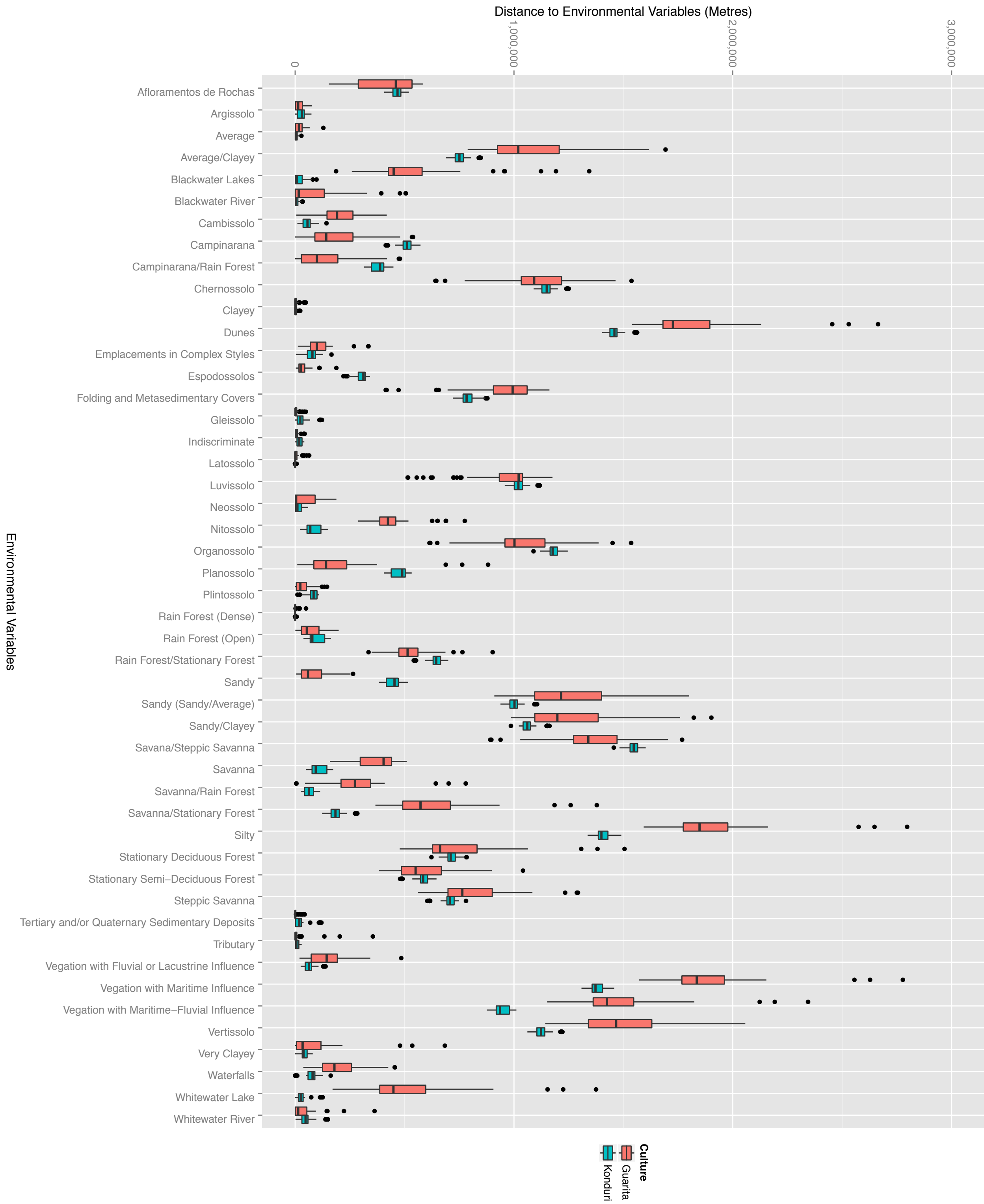


Figure 4.7: Overview of the distances to environmental variables

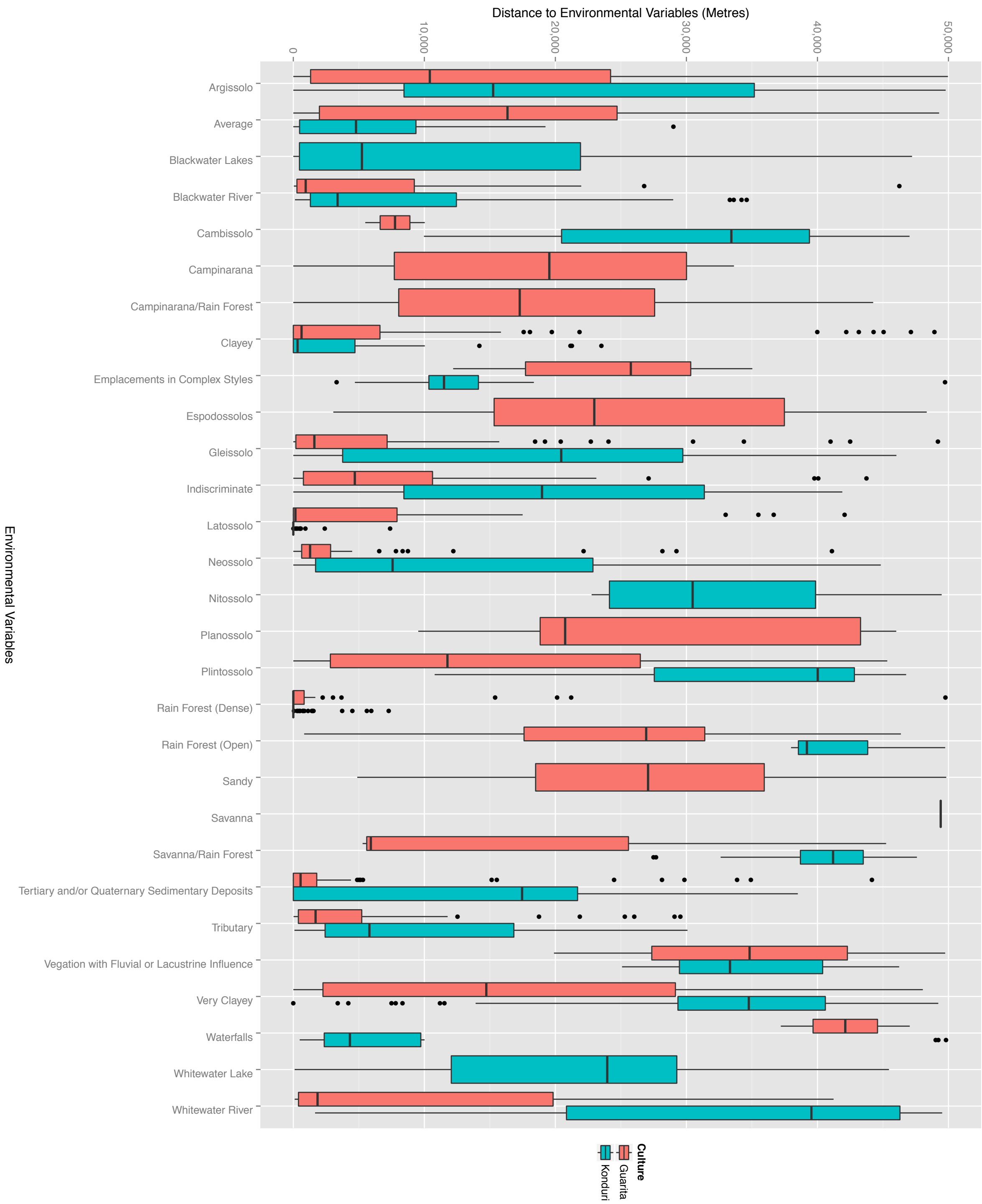


Figure 4.8: Overview of the environmental variables within 50 kilometres

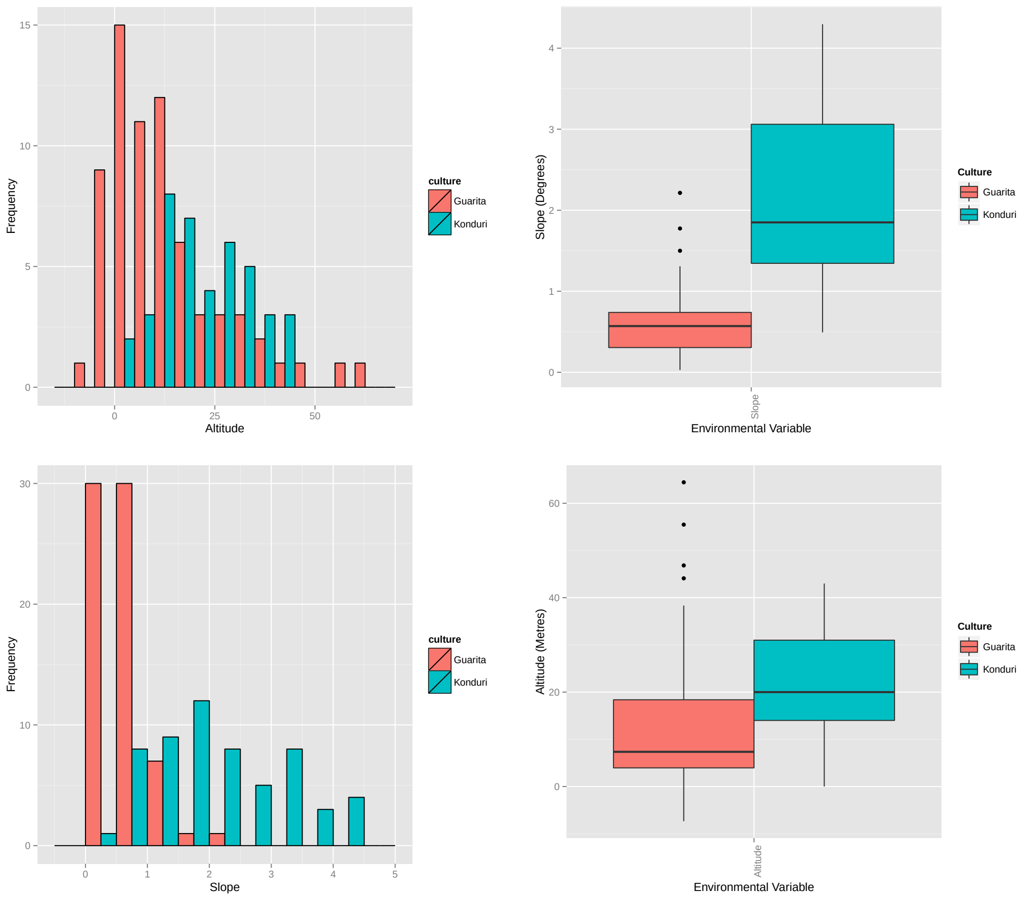


Figure 4.9: Overview of the distribution of the altitude above water level (top) and slope (bottom) of the excavation sites

ones from the Konduri culture. A major difference is the absence of lakes close to settlements of the Guarita culture. 25 of the initial 42 variables are used for further analysis. Both cultures seem to prefer settlements which are located above the water level of the rivers. Differences can be identified according to the height. Whereas the known sites from the Guarita culture mainly are found 5 to 20 metres above water level settlements from the Konduri culture are mainly located 15 to 30 metres above water level. These first simple plots already allow first drawbacks about the settlements and facilitate the comparison of various cultures. Some of the selected distance measures correlate. The figures 4.10 and 4.11 visualize the correlation between the environmental variables. Red colors indicate a negative correlation whereas blue colors indicate a positive correlation whereas the color intensity and size are proportional to the correlation coefficients. Some soil types and conditions are correlated (or negatively correlated) as well as soil types and/or conditions with the vegetation and water/river type. Additionally some water types are correlated (e.g. whitewater lakes are always located next to white water rivers). This is not surprising because of the interdependencies of these variables. Vegetation depends on the availability of water as well as on the soil and climate. The soil condition, respectively the soil type is influenced by the availability of water as well as the availability of nutrients.

4.4.1 Determining the importance of environmental variables

The figures above (figure 4.7 and figure 4.9) allow the elimination of some environmental parameters. In order to further identify the importance of each variable, the MEM approach is used. All environmental parameters within 50 kilometres, precipitation as well as the slope and altitude above water level are used as input for the MEM. Some properties need to be set in advance before running the model. The dataset provides a collection of presence data but does not include reliable absence information. In order to create so called pseudo absences, MaxEnt randomly selects background samples. They are needed in order to estimate the distribution of environmental parameters across the model extent. The extent is limited to the extent of the excavation sites plus a 100 kilometre buffer to ensure that the background sample has the same bias like the presence locations. In case of the Guarita culture a buffered convex hull is used to assure a coherent area. In order to test the model variability, the model can perform replicate runs which can be compared. This parameter is set to ten in order to determine the variability and get a more reliable (average over ten runs) output.

In order to evaluate the model performance, several output files are created by the

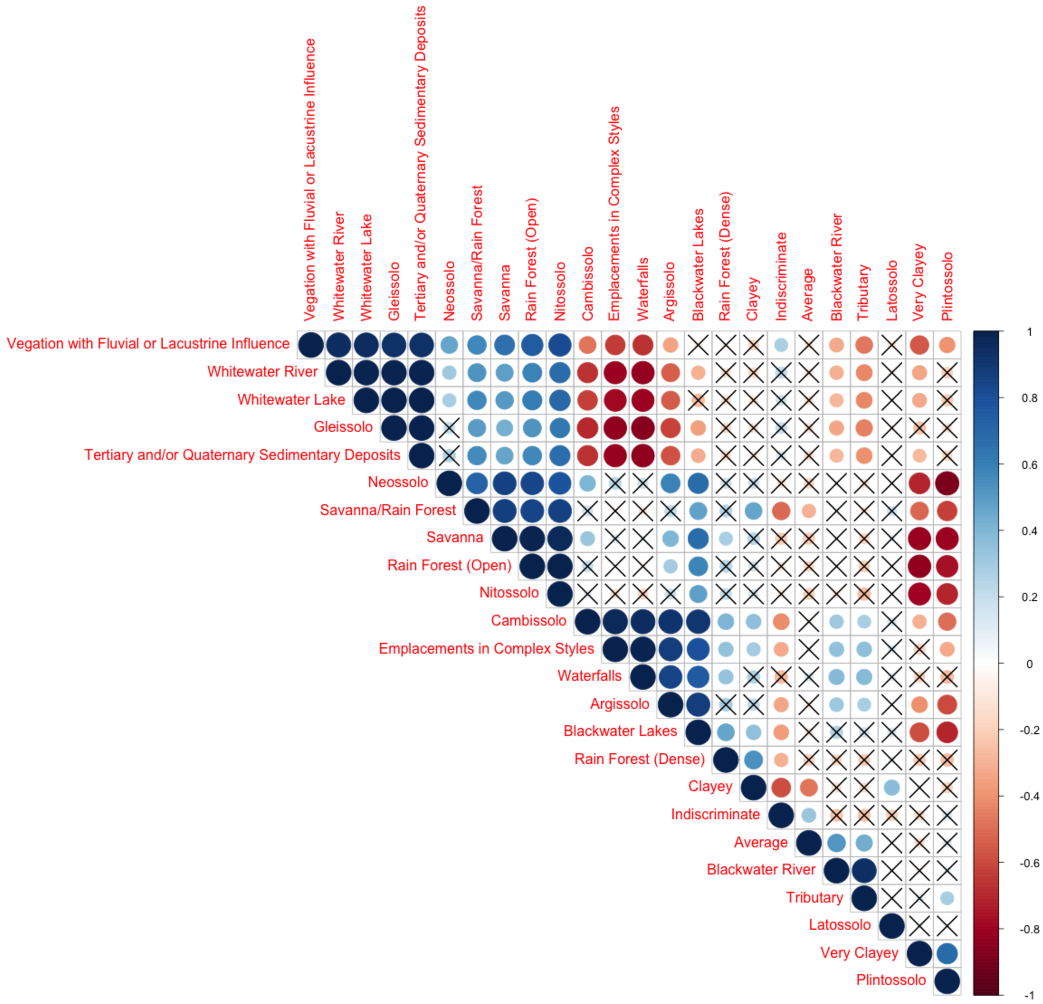


Figure 4.10: Overview of the correlations between the distance measures (Konduri)

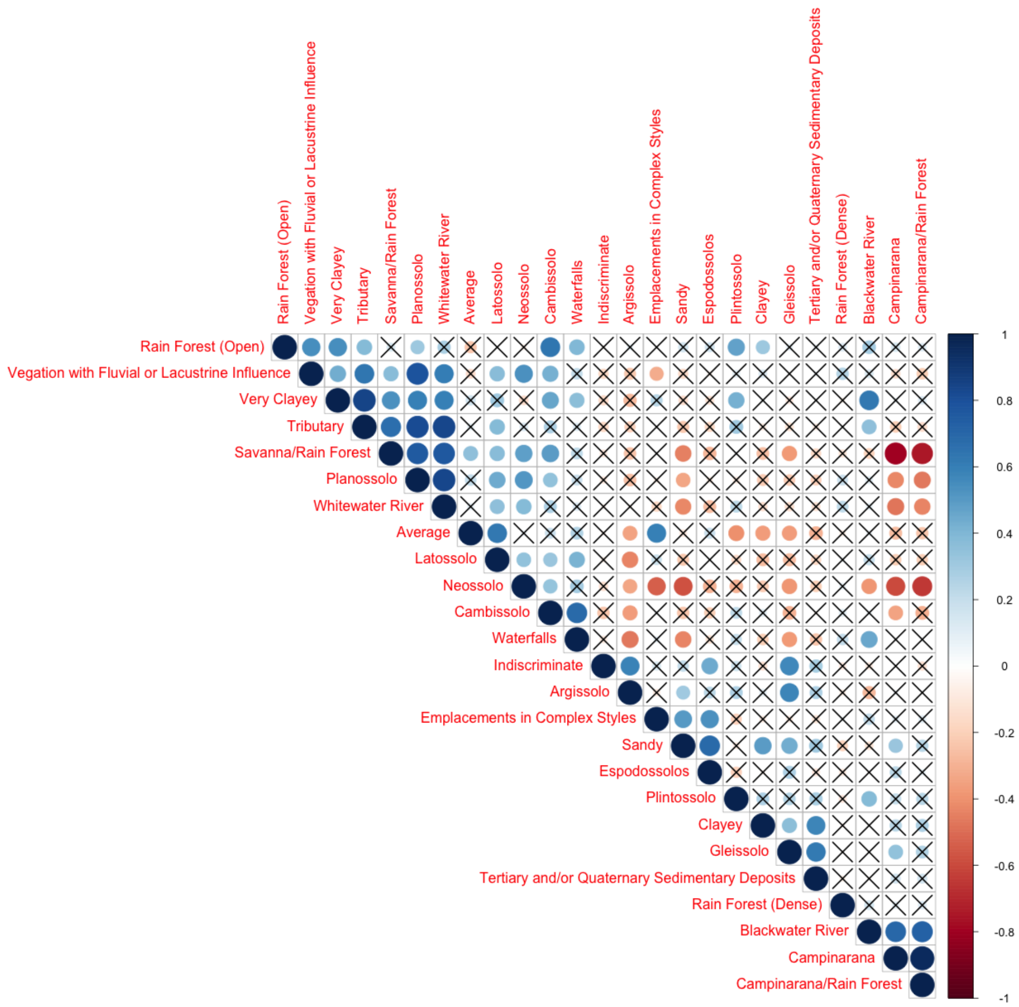


Figure 4.11: Overview of the correlations between the distance measures (Guarita)

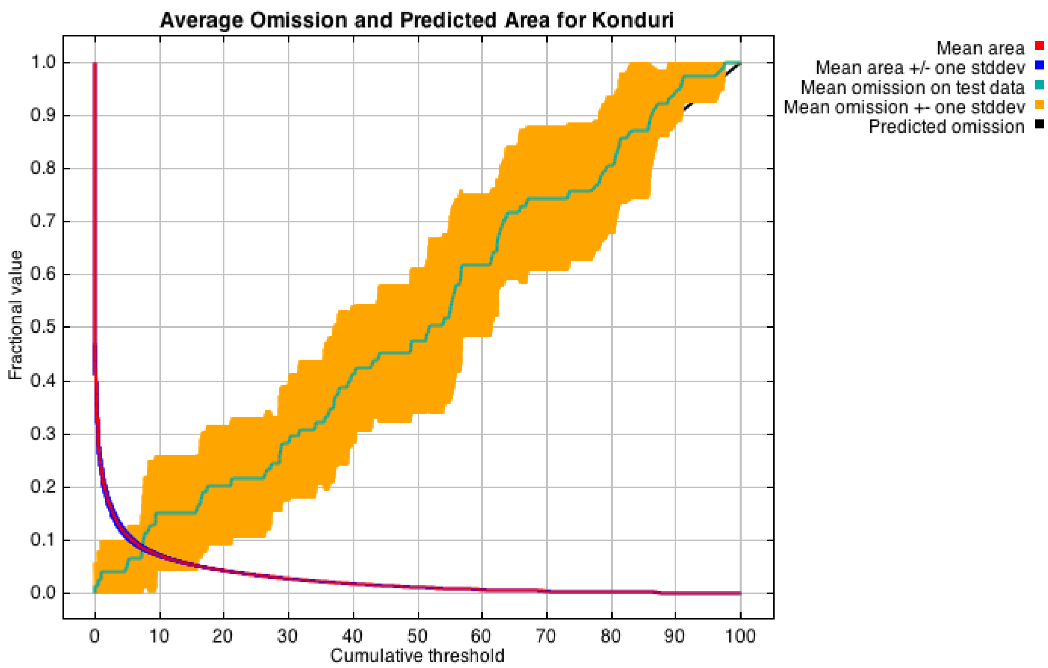


Figure 4.12: The average omission and predicted area for the Konduri culture

MaxEnt program. The potential sample bias is analysed using the omission rate. The graph depicts the relationship between predicted habitat suitability values and the occurrence sample. Thus, it shows the number of locations that do not occur in a suitable habitat. Unless the prediction is biased, the omission rate should be close to the predicted omission, hence a 1 : 1 relationship. The figures 4.12 and 4.13 show the omission rate for the model runs of both the Konduri and the Guarita model. Whereas the mean omission (green line) is close to the predicted omission (black line, which is overlaid in figure 4.12) for the Konduri model, the mean omission is above the predicted omission for the Guarita model. The model performance is tested using the area under the receiver operating characteristic (ROC) data – or area under the curve (AUC) data) – value. It compares the predicted occurrences to a random selection of points. A mean value of 0.5 indicates that the performance is not better than random whereas a mean value close to 1.0 suggest a good model performance. This allows the comparison of multiple models (Hanley and McNeil 1982). As shown in the figures 4.14 and 4.15 the average AUC value for the Konduri model is 0.967 with a standard deviation of 0.018 which indicates a good model performance whereas the performance for the Guarita model run is slightly worse with an average AUC value of 0.931 and a standard deviation of 0.069. How much each variable effects the model is determined using response curves. They are derived in two different ways:

- only varying one environmental variable whereas all other environmental

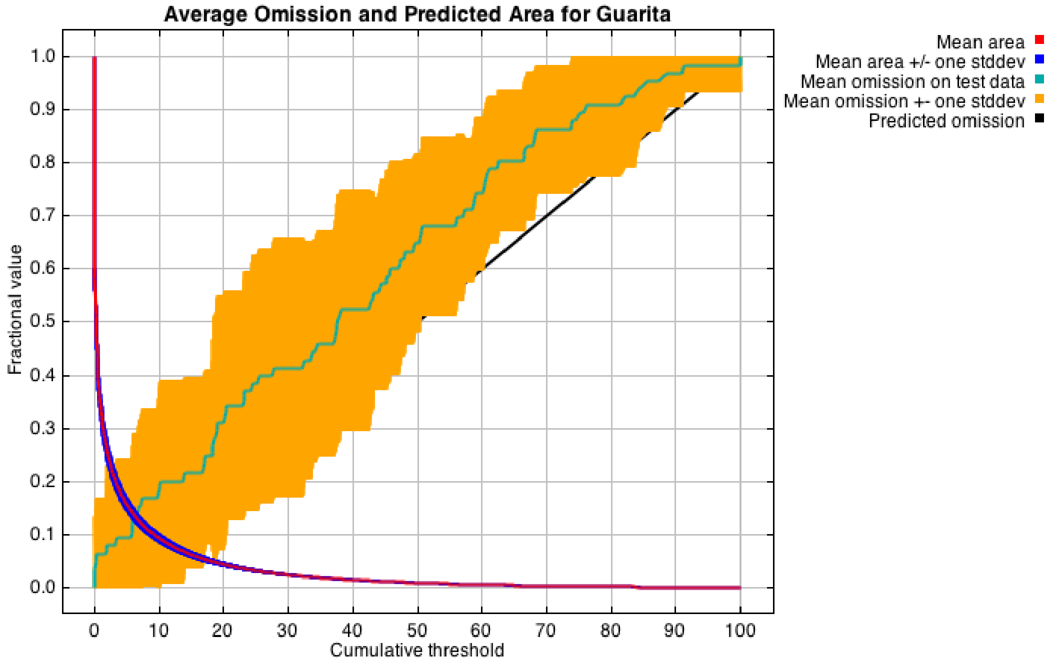


Figure 4.13: The average omission and predicted area for the Guarita culture

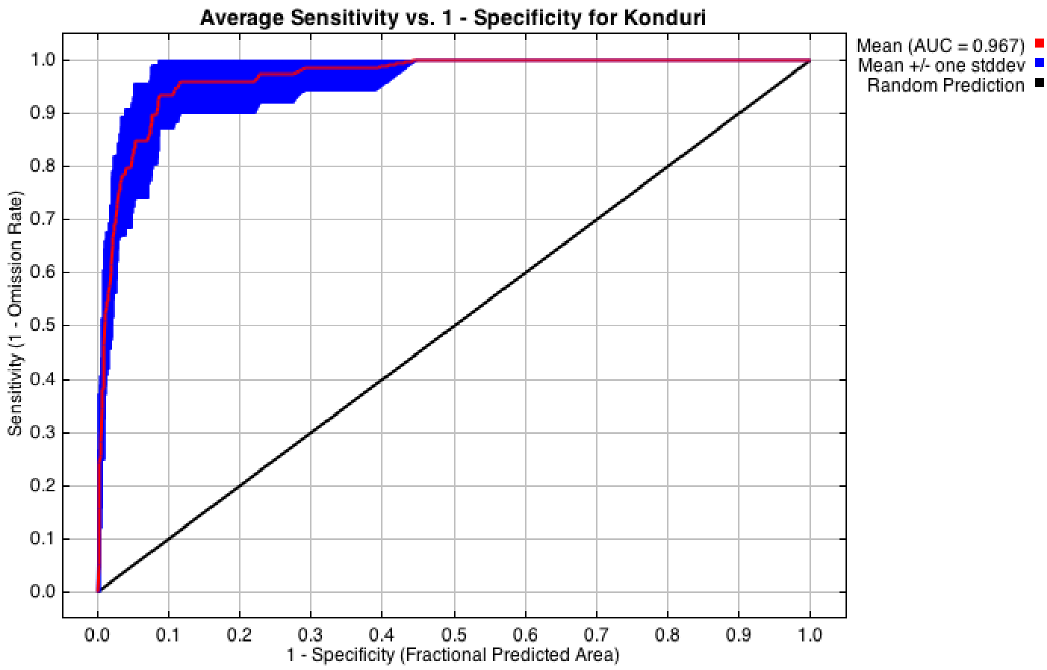


Figure 4.14: The graph of the ROC for the Konduri model

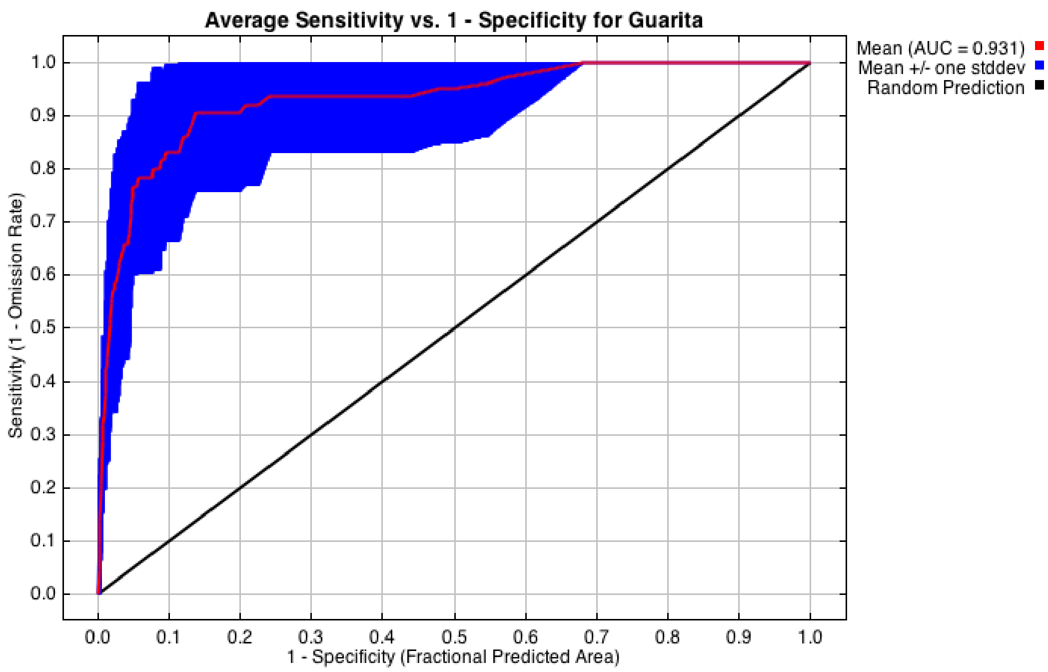


Figure 4.15: The graph of the ROC for the Guarita model

variables keep their mean sample value

- creating a model using only the corresponding variable

The marginal response curves are difficult to interpret if the correlation between variables is high. Due to the high correlation of some of the environmental variables (as shown in figures 4.10 and 4.11) the second approach is used. The response curves show its value (in most cases the distance) of the environmental variable on the x axis and probability of presence (logistic output) on the y axis. The results of the first approach can be found in the appendix (figures H.16 and 4.16). Two tables with the calculated importance weights are displayed, namely the percent contribution and the permutation importance. Both provide an estimate of the relative contribution of the environmental parameters to the model. In order to determine the percent contribution the increase in gain is added to the contribution of the corresponding variable for each iteration. The permutation in contrast is determined by permuting random values for training presence and background data. The resulting drop in the training AUC is used for the permutation importance. The displayed values are averages over the various model runs (see figures 4.18 and 4.19 for the results of the models). As with the response curves the percent contribution is difficult to interpret if the correlation is high. Based on the partly strong correlations the permutation importance is used for determining the variable weight. Another output is a suitability surface which considers the locations of the former settlements

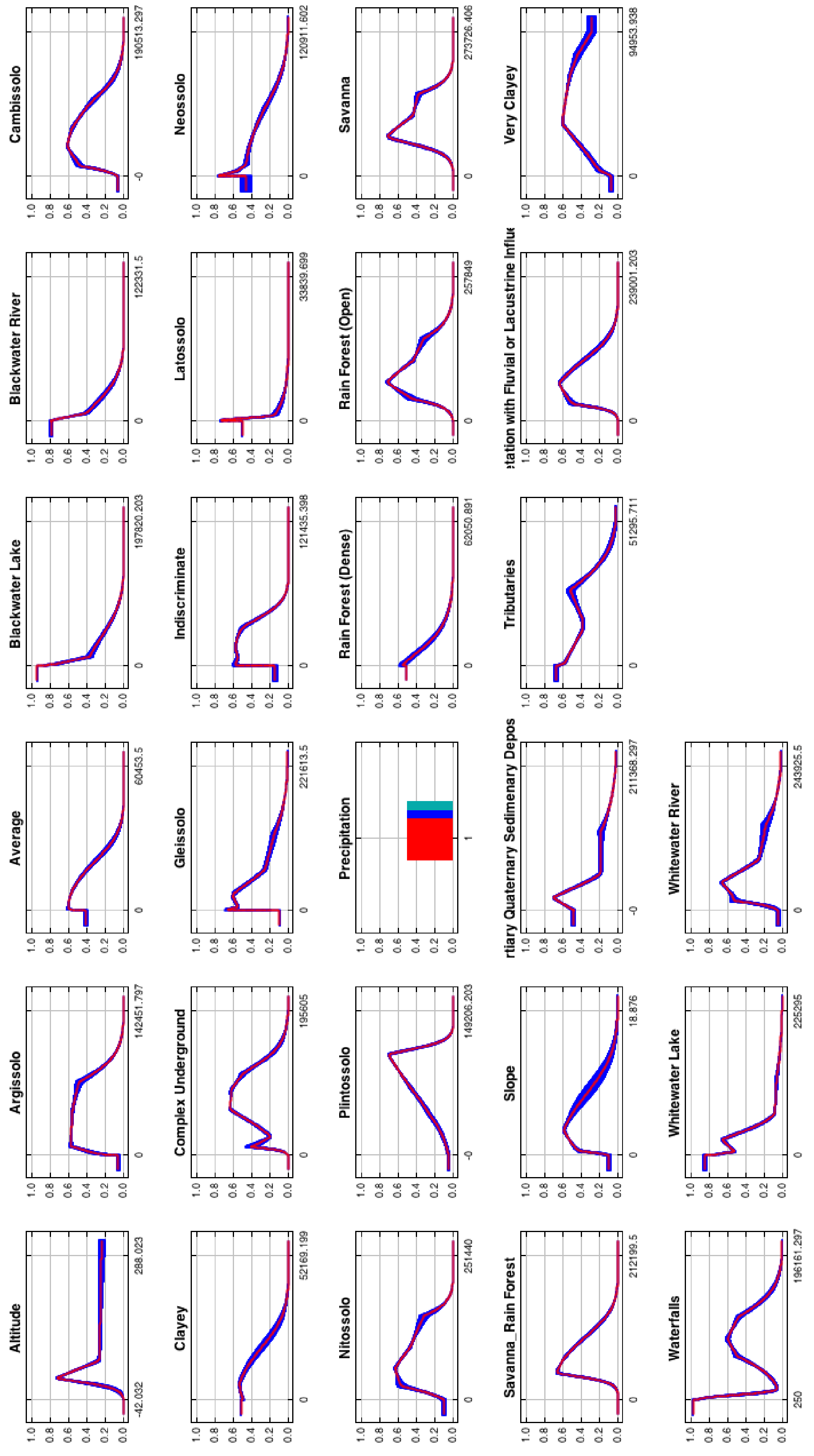


Figure 4.16: The response curves for all environmental variables derived using only the corresponding variable as model input. This is the result for the Konduri model.

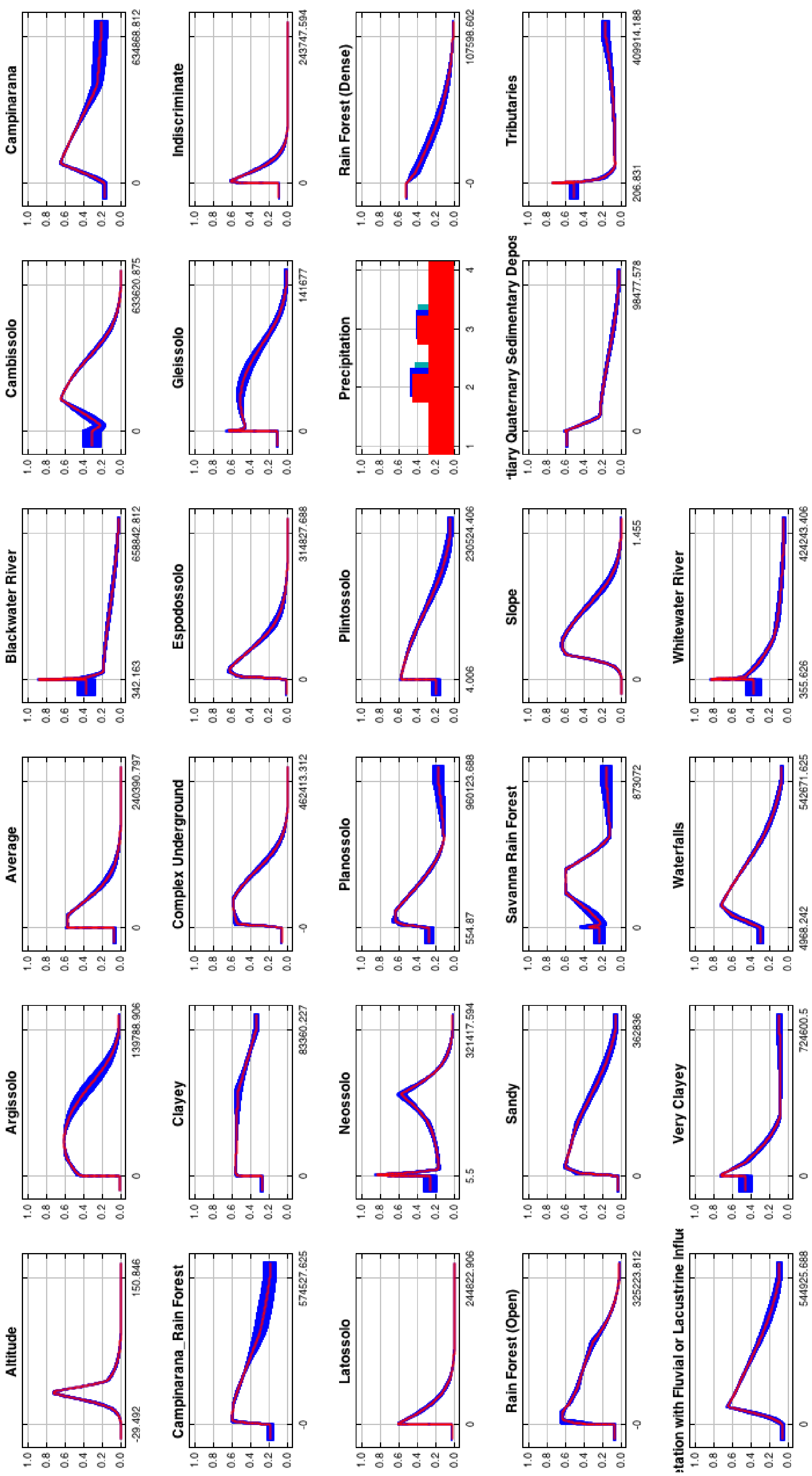


Figure 4.17: The response curves for all environmental variables derived using only the corresponding variable as model input. This is the result for the Guarita model.

Variable	Percent contribution	Permutation importance
Blackwater River	38.1	6.9
Whitewater River	16.6	4.9
Tributaries	6.6	2.4
Indiscriminate	5.7	27.6
Latossolo	5.6	21.6
Espodossolo	5.5	7.2
Precipitation	4.6	0.9
Neossolo	4.3	1.2
Argissolo	2.7	2.2
Campinarana_Rain Forest	1.9	0.4
Gleissolo	1.5	0.8
Average	1.2	1.9
Cambissolo	0.8	2.5
Vegetation with Fluvial or Lacustrine Influence	0.7	0.5
Slope	0.7	1.9
Savanna Rain Forest	0.6	2.7
Altitude	0.6	10.3
Complex Underground	0.5	1
Clayey	0.5	0.6
Tertiary Quaternary Sedimentary Deposits	0.3	0.3
Sandy	0.3	0.2
Very Clayey	0.2	0.1
Plintossolo	0.2	0.1
Planossolo	0.1	0.5
Campinarana	0.1	0.3
Waterfalls	0	0.1
Rain Forest (Open)	0	0.9
Rain Forest (Dense)	0	0.2

Figure 4.18: The variable contributions as percent contribution as well as permutation importance for the Konduri model

Variable	Percent contribution	Permutation importance
Blackwater Lake	32.1	26.4
Whitewater Lake	21.7	16
Latossolo	11.5	14.6
Blackwater River	8.9	1.8
Argissolo	4.7	3.7
Very Clayey	3.1	8.6
Savanna	2.9	0
Nitossolo	2.5	0.9
Slope	2.4	2.3
Waterfalls	2.1	1.1
Altitude	2	3.3
Average	1.3	3.1
Tributaries	0.8	0.5
Gleissolo	0.7	0.7
Indiscriminate	0.7	8.7
Complex Underground	0.7	0
Tertiary Quaternary Sedimenary Deposits	0.6	0.2
Neossolo	0.5	3
Vegetation with Fluvial or Lacustrine Influence	0.4	0.4
Clayey	0.3	1.9
Savanna_Rain Forest	0.1	0.2
Whitewater River	0.1	0.2
Cambissolo	0	0.1
Plintossolo	0	0
Precipitation	0	0
Rain Forest (Dense)	0	0
Rain Forest (Open)	0	2.3

Figure 4.19: The variable contributions as percent contribution as well as permutation importance for the Guarita model

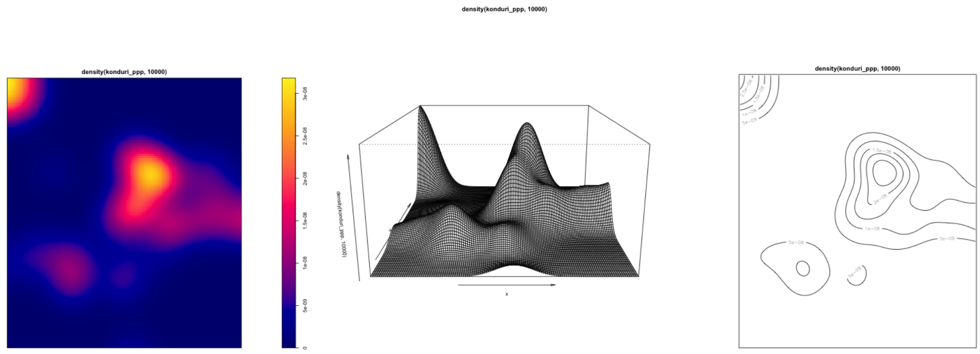


Figure 4.20: Three different kinds of density plots (2D density plot on the left, 3D density plot in the middle and contour lines on the right) for the Konduri culture

as well as the input parameters. Additionally, summary rasters with the standard deviation, minimum, maximum and median are provided. Further outputs are the results of the jackknife test in order to test the variable importance. Jackknife is a resampling technique to identify bias and variance in the model. Two properties are depicted in the figure, namely the training gain without the corresponding variable and with only the corresponding variable. Whereas the former indicates the usefulness of the information itself, the latter describes which variable has the most information which is not present in the other variables. All outputs can be found in the appendix (chapters H.1 and H.2).

4.4.2 Identifying the settlement types and their predominant environmental variables

The results from the maximum entropy modelling process are derived using all locations of the observed cultures (namely Konduri or Guarita). Thus the results such as the suitability surface are general and do not distinguish between functional settlement types. Due to that only the variable importance is used for further input to built the suitability surface. Before the suitability calculation can be performed, a clustering process is needed in order to determine the various settlement types. The clustering process requires a point pattern analysis to assure a clustered pattern. If a random or regularly dispersed pattern is identified, a clustering analysis is misleading. Consequently, it needs to be clarified if the locations show a clustered pattern.

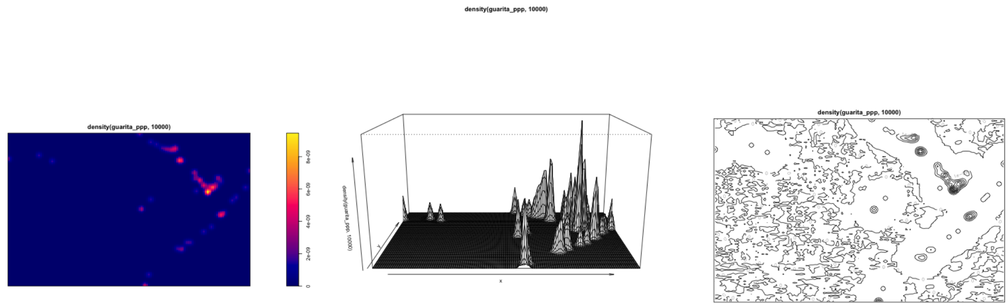


Figure 4.21: Three different kinds of density plots (2D density plot on the left, 3D density plot in the middle and contour lines on the right) for the Guarita culture

4.4.2.1 Check for Clustered Patterns

A first impression of the distribution can be gained by looking at a density plot. The figures 4.20 and 4.21 show three different types of density plots, namely a 2D heatmap, a 3D surface as well as a contour lines plot. The plots suggest that a clustered pattern is present. A more detailed analysis of the settlement pattern can be derived using the G-, F- and K-function. The graphs in the figures 4.22, 4.23 and 4.24 respectively 4.25, 4.26 and 4.27 show the results of the three functions. The black line $< function_{l}etter >_{obs} (r)$ equals the variable $< \widehat{function_{l}etter} >(r)$ which is mentioned in chapter 3.2.1. The red line indicates the theoretical distribution and the grey area indicates the simulated envelope. These figures indicate a strongly clustered pattern for both observed cultures. An observed distribution function $\hat{G}(r) > G(r)$ ($\hat{G}(r)$ equals $G_{obs}(r)$ in figure 4.22) indicates a clustered pattern. In contrast, the F-function indicates a clustered pattern if $\hat{F}(r) < F(r)$ ($\hat{F}(r)$ equals $F_{obs}(r)$ in figure 4.23). The K-function allows to identify pattern changes at different scales. This is useful if a point pattern is random in small scale but clustered in large scale. Similar to the G-function, a value $\hat{K}(r) > K(r)$ ($\hat{K}(r)$ equals $K_{obs}(r)$ in figure 4.24) indicates a clustered pattern. If a pattern change occurs, the relationship between $\hat{K}(r)$ and $K(r)$ would vary. Additionally, a Monte Carlo simulation was implemented. Figure 4.28 shows an exemplary result of the Monte Carlo simulation with the increasing number of excavation sites on the x-axis and the distance towards the environmental parameter on the y-axis. The black line depicts the lowest distance between all excavation sites and blackwater lakes. The same amount of points as observed excavation sites (namely 88 random points in the case of Konduri and 69 in the cas of Guarita) were randomly created in the same spatial extent and the minimum distance towards blackwater lakes is measured. This process was

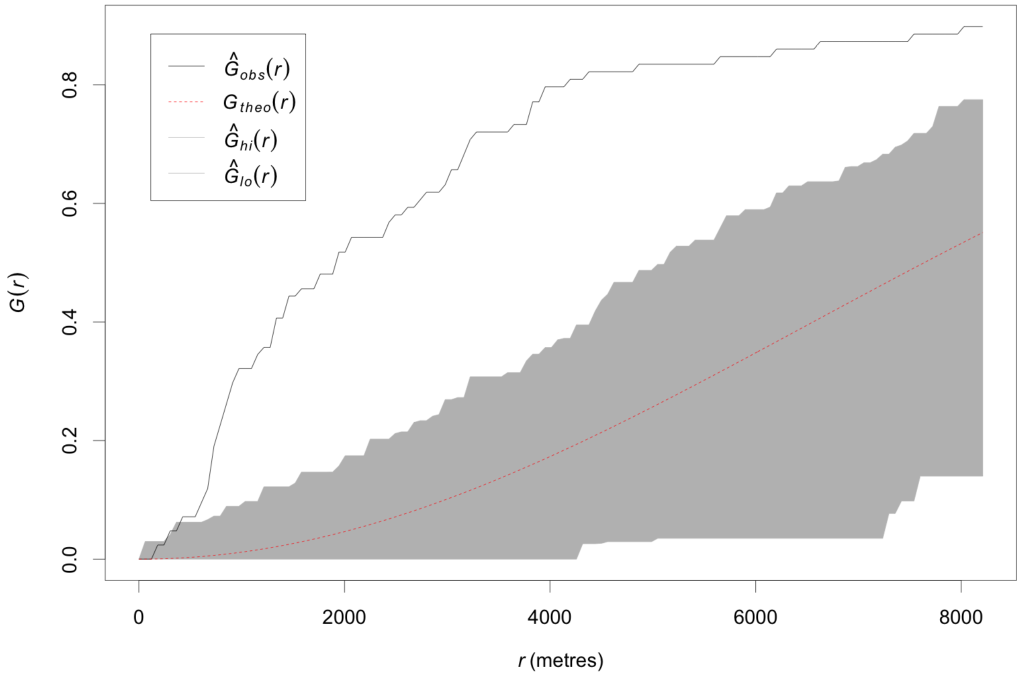


Figure 4.22: G-function for Konduri settlements

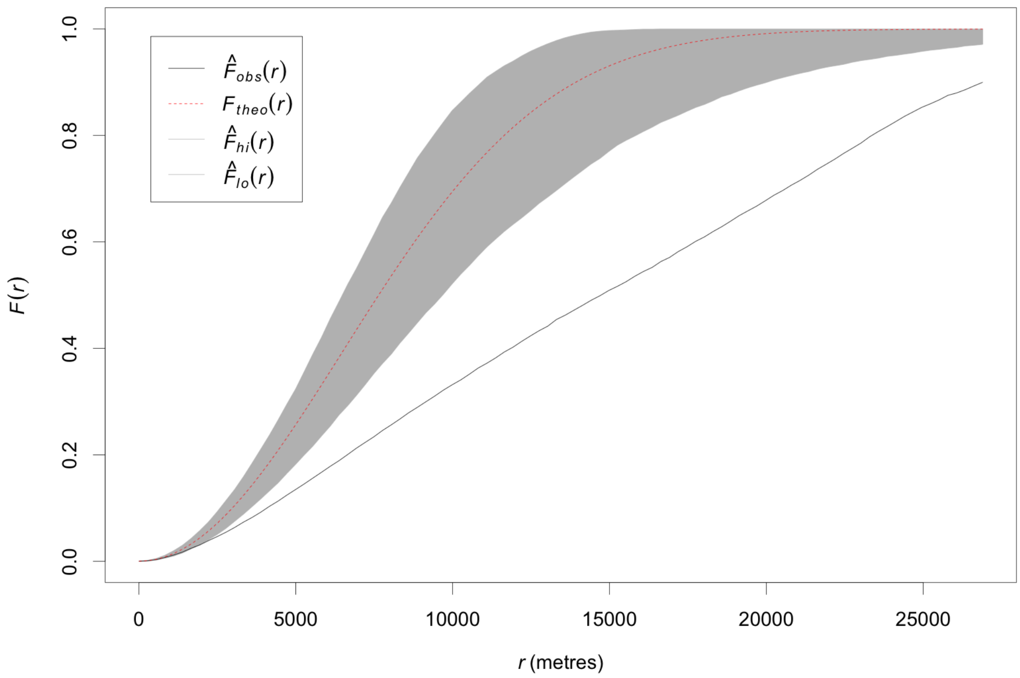


Figure 4.23: F-function for Konduri settlements

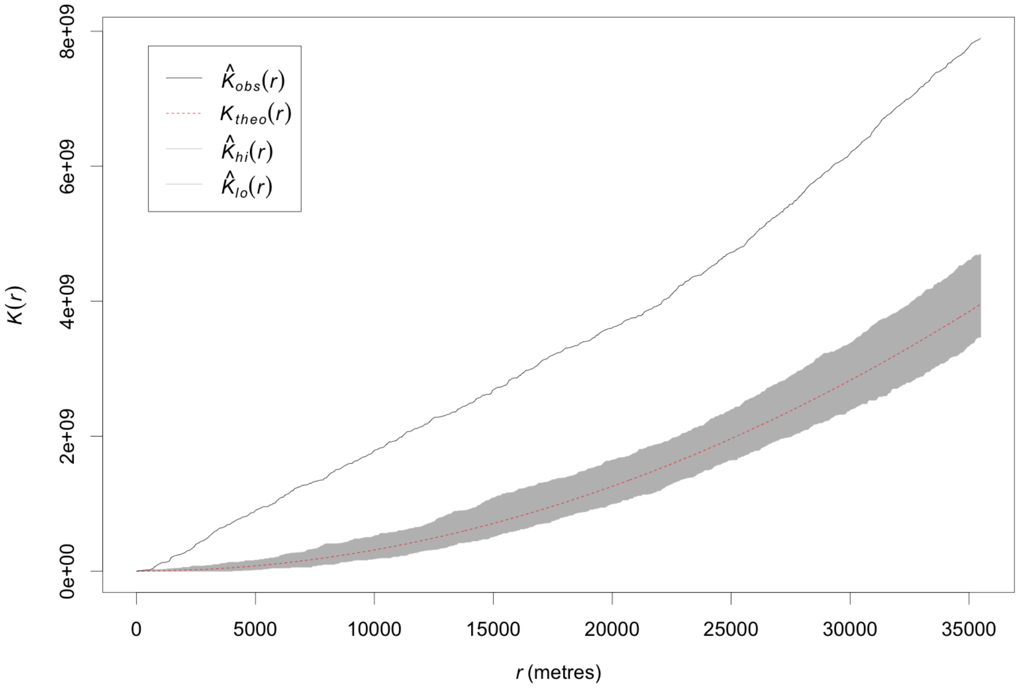


Figure 4.24: K-function for Konduri settlements

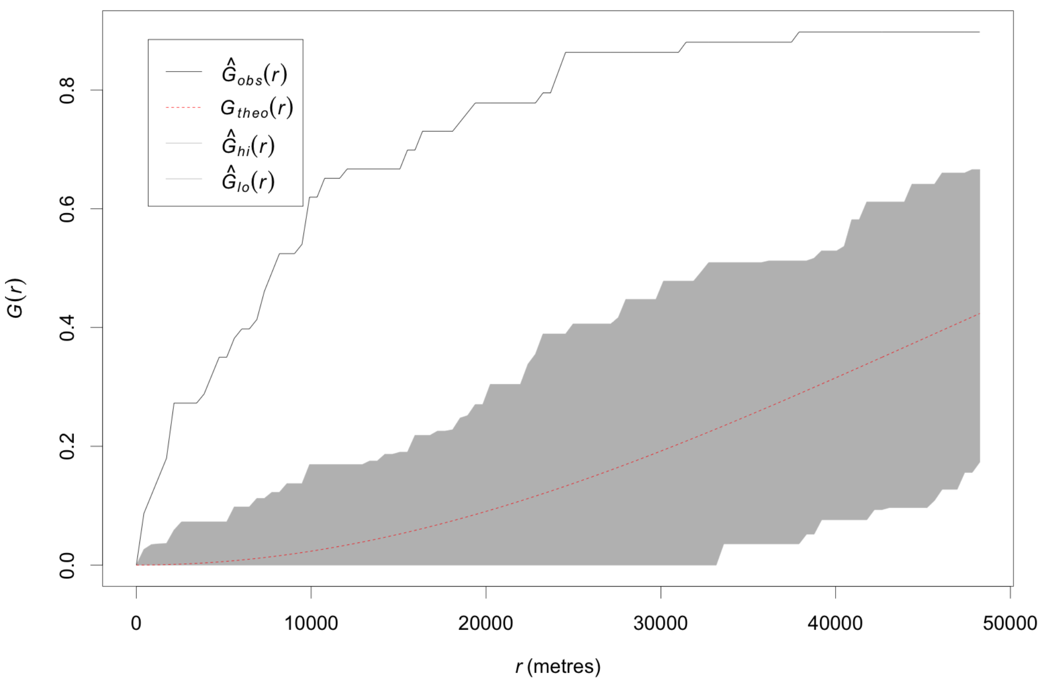


Figure 4.25: G-function for Guarita settlements

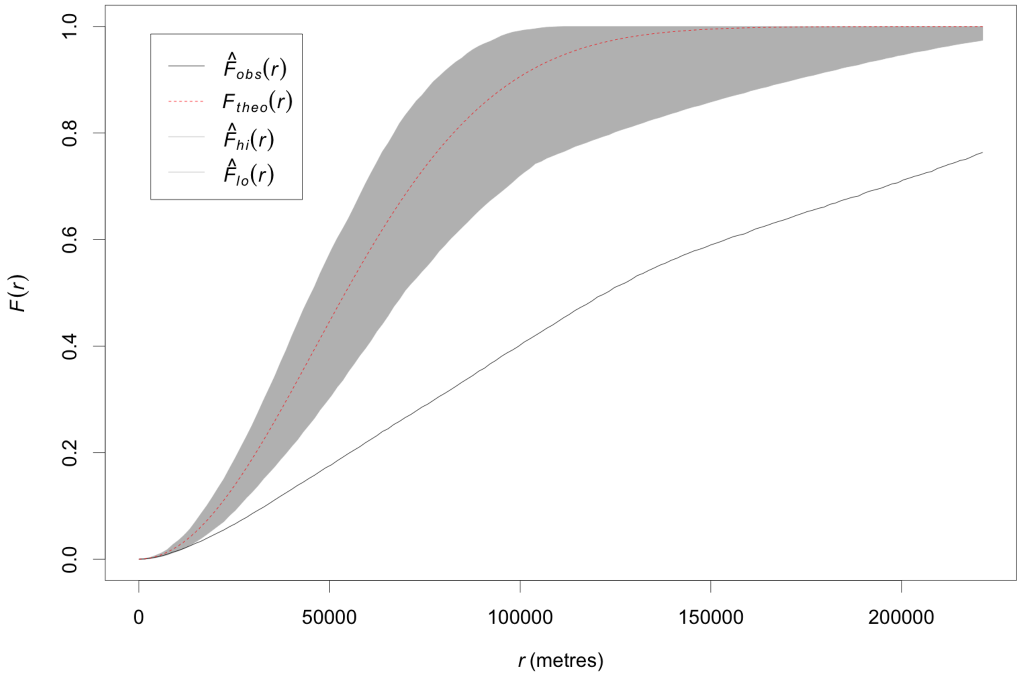


Figure 4.26: F-function for Guarita settlements

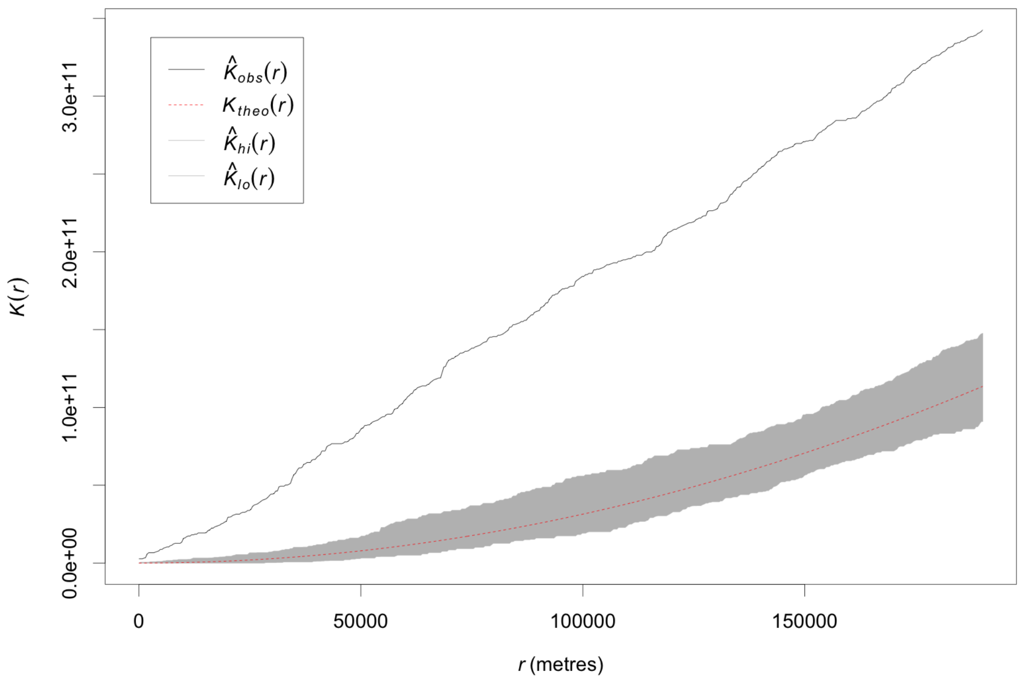


Figure 4.27: K-function for Guarita settlements

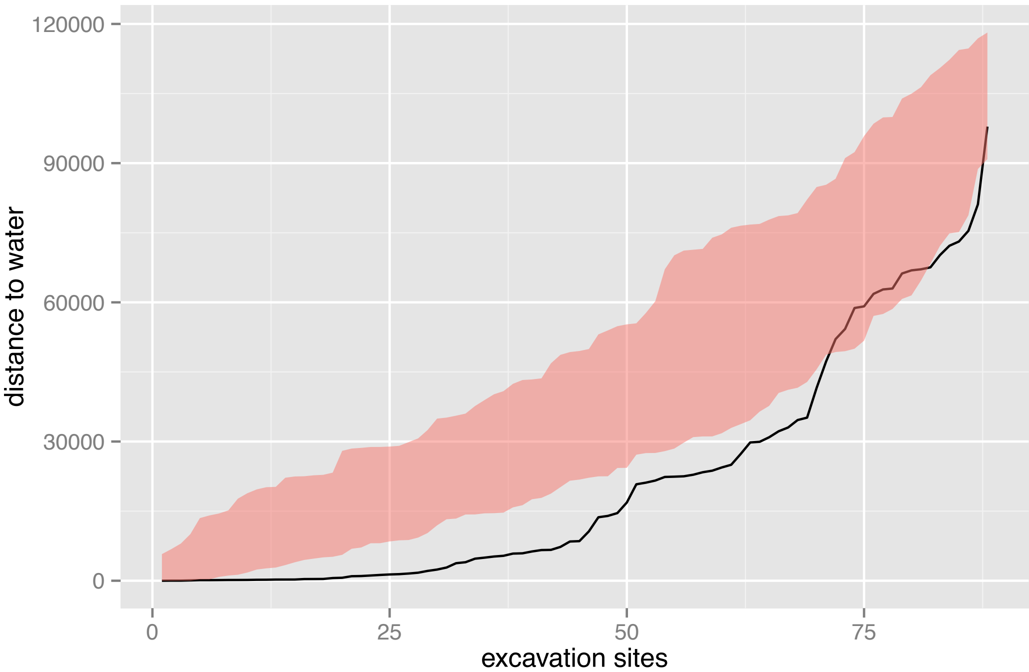


Figure 4.28: Results of the Monte Carlo simulation. The excavation sites are shown on the x-axis and the distance is shown on the y-axis. The black line indicates the distance (sorted in ascending order) from the excavation sites to the environmental parameter whereas the envelope depicts the minimum and maximum distances (sorted in ascending order) of the randomly dispersed points. This example shows the distance of black-water rivers to the excavation sites of the Konduri culture.

repeated several times – 1000 times in total in this case.The red envelope shows the minimum and maximum measures of all Monte Carlo runs (the results for all environmental parameters can be found in the appendix J). If the black line is within the red envelope, the measured distances do not differ from the randomly created location measures. In the case of blackwater lakes it can clearly be seen, that the black line is almost completely below the red envelope. That means that most of the excavation sites are closer to blackwater lakes than randomly dispersed locations in the same spatial extent. Therefore, the Monte Carlo simulation provides detailed information about the environmental preferences of former settlements. The Monte Carlo simulation is applied to all excavation sites. Those results can be seen as an additional gain of knowledge but – due to the absence of spatial information – have no influence on the further process.

4.4.2.2 Determine the number of different settlement types

As assumed, the settlement functionalities (at least partly) rely on the environmental surroundings. Thus, the number of clusters – with the environmental data as input data – determines the number of settlement functions of the observed cultures. In order to identify the clusters, all variables with an permutation importance value > 0 are used. As explained in chapter 3.2.3 a table provides the outer ($max_{near_{outer}}$) and inner ($max_{near_{inner}}$) maximum nearness boundary. This allows to set individual values for various environmental variables as well as cultures. A value within that boundary is chosen randomly for each environmental variable and used for the cluster analysis. This process is repeated many times (1000 times in this case), which leads to a collection of different clustering results with different maximum nearness values. The k-means clustering algorithm needs a predefined number of clusters which can be determined in several ways and varies according to the used approach. In order to develop a preferably automated workflow, it is important to choose a method which needs no interpretation from the user (such as the elbow method, which uses a graph to facilitate the user's decision). As mentioned in chapter 3.2.3, the consensus clustering approach was initially developed to determine the number of classes unsupervised (Monti et al. 2003). This is used to avoid additional methods to determine the number of clusters. The consensus clustering value can be used for the comparison of different cluster results. Thus the value cannot only be used in order to estimate the best cluster solution but also to compare different cluster runs with varying cluster numbers k. The consensus cluster values for the varying cluster numbers are shown in figure 4.29. As can be seen, the optimal cluster number for both cultures is five.

4.4.2.3 Calculation of Settlement Function Related Environmental Influences

After determining the optimal number of clusters, the identification of environmental influences can be made based on the settlement function. Therefore, the consensus clustering process is performed which is set up with the following distance settings. All environmental variables except the distance to waterfalls is defined using 500 metres as $max_{near_{inner}}$ and 5000 metres as $max_{near_{outer}}$. The $max_{near_{outer}}$ is extended to 15000 metres for the distance to waterfalls. This values are used to present the methodology bur can be set individually. The result for five cluster numbers and 1000 cluster runs can be seen in the following figures (figures 4.30 – 4.39). The figures 4.30, 4.31, 4.32 and 4.33 show the item consensus values for each cluster and excavation site. The x-axis shows the excavation site id in all four plots. The difference between the figures 4.30, 4.31 and 4.32, 4.33 lie in the different scale

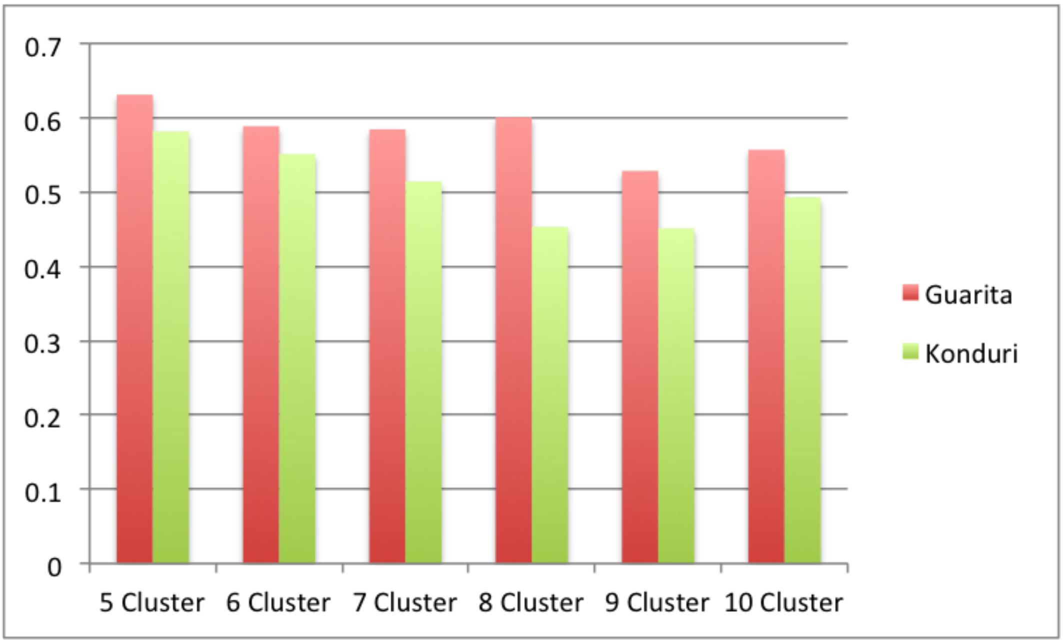


Figure 4.29: Comparison of various cluster numbers using the mean consensus cluster value

on the y-axis. Whereas the values in the figures 4.30 and 4.31 are the cumulative item consensus values, the values of 4.32 and 4.33 are normalized to fit 100 percent. As can be seen, some excavation sites have at least one cluster with a high item consensus value (e.g. the items on the left of figure 4.30 which are assigned to cluster two (red) and have an item consensus value ≥ 0.8). Other excavation sites do not have a high item consensus value in any of the clusters (e.g. excavation site number 42 in the middle of the figure 4.30). These values have influence on the cluster consensus values. Figures 4.34 and 4.35 show the cluster consensus results with the cluster numbers on the x-axis and the cluster consensus values on the y-axis. If many items with relatively low item consensus values are assigned to one cluster, the cluster consensus value will also be low (e.g. cluster 3 in figure 4.34). The figures 4.36 and 4.37 give additional information about the consensus clustering result. These graphics are heatmaps of the consensus matrix m_{cons} whereas the color intensity ranges from dark blue (always clustered together) to white (never clustered together). The dendrogram on top shows the clustering hierarchy and the color coded values below show the cluster result for a cluster number of 5. The results can be used to further define the environmental conditions of specific functional settlement pattern. The consensus clustering process returns item consensus values which defines the most likely cluster and its consensus value for each excavation site. This cluster solution can be applied to the observed excavation sites which is shown in the figures 4.38 and 4.39. An excerpt of the output can be seen in table 4.1.

In the next step the item consensus values for each excavation site and its optimum cluster are extracted and used for further analysis. In a second step, the measured distance values are used to determine the environmental properties of the settlement type. Therefore, the item consensus values are normalized so that the values range between 1 and 0 (as shown in 4.2). The distance measures are used as values of truth in a fuzzy membership function which is computed for each environmental variable and cluster (as shown in figure 4.40). The results of the clustering process show that only a small set of environmental variables is influencing the settlement type. This can be used to define pattern specific properties. Cluster 4 is built based on the distances towards blackwater rivers, the soil Latossolo and the soil type average. In the case of Konduri settlements the number of predominant environmental variables (item consensus value > 0.8) are listed in table 4.3. They can be reduced to 11 out of the initial 42 input variables. In the case of Guarita 15 out of 42 variables can be identified.

4.4.2.4 Calculation of the Suitability Surface and Territory

The fuzzy membership function in combination with the optimal cluster solution for each excavation site allows to determine a settlement function related suitability surface. This is done by applying the distance and item consensus values for each environmental parameter of the same cluster to the environmental variables. The result are multiple buffers which can than be used to calculate a settlement type (i.e. cluster) related suitability surface. Therefore, the various resulting multiple buffers are combined cluster-wise using the permutation importance from the MEM. Subsequently, the assigned suitability values are normalized between 1 and 100. The resulting suitability surfaces S_{clust} (as shown in figure 4.42 – all surfaces can be found in the appendix K) can be seen as a surface of settlement likelihood for the observed culture and functional settlement pattern.

S_{clust} can then be used to calculate the territory of the observed culture. The $Dist_{max}$ value is determined by the maximum distance towards an influencing environmental variable. E.g. cluster 3 of the Konduri culture is described by the distances toward several variables (as shown in table 4.3), whereas the distance of almost 10.000 metres towards the waterfalls is the maximum distance, therefore $Dist_{max}$ for cluster 3 is set to 10.000 metres. The method has the advantage that the $Dist_{max}$ value can be individually set based on the settlement type rather than setting the same $Dist_{max}$ value for all settlements.

The result for the functional settlement pattern of cluster 1 of the Guarita culture is shown in figure 4.43 (the results of all territories can be found in the appendix L).

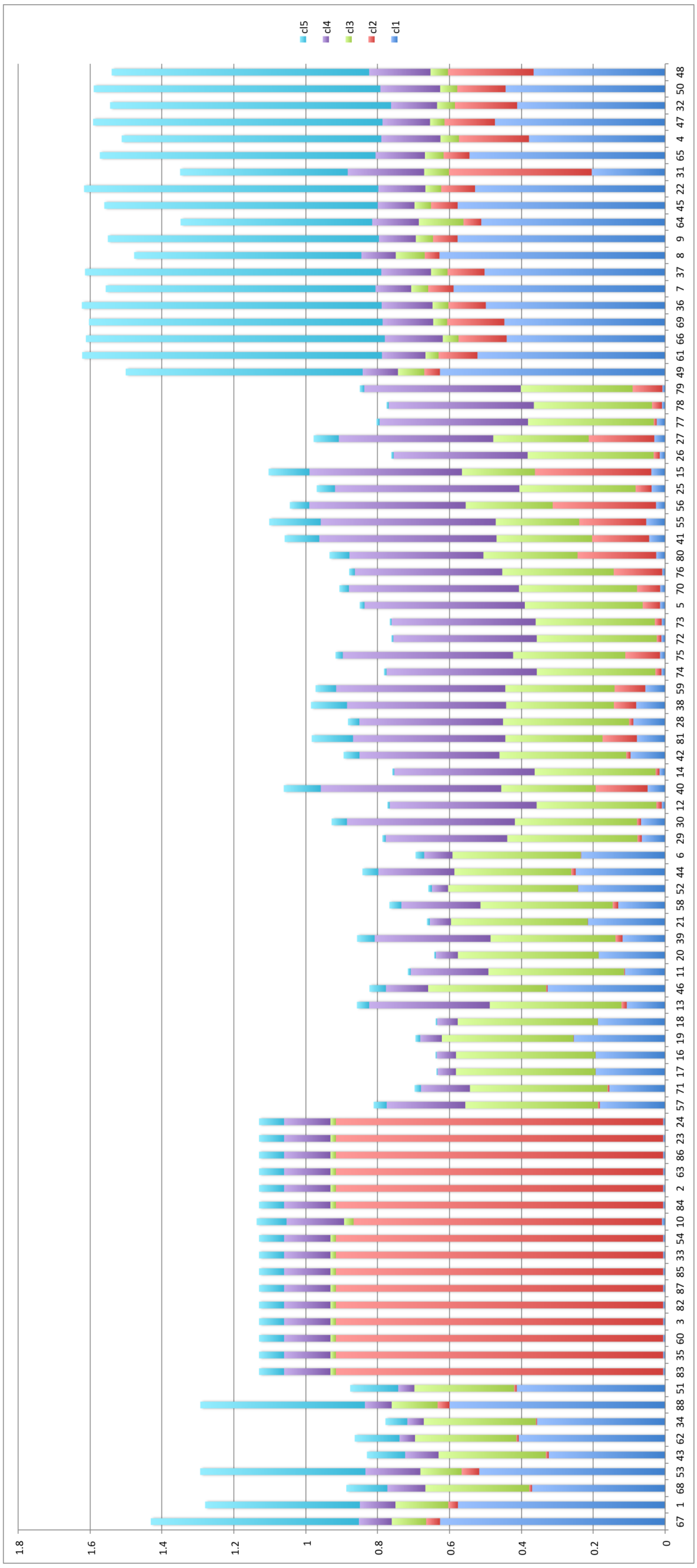


Figure 4.30: The item consensus values for each excavation site of the Konduri culture and cluster. The higher the bar of one color the higher the consensus value for that cluster for a specific excavation site.

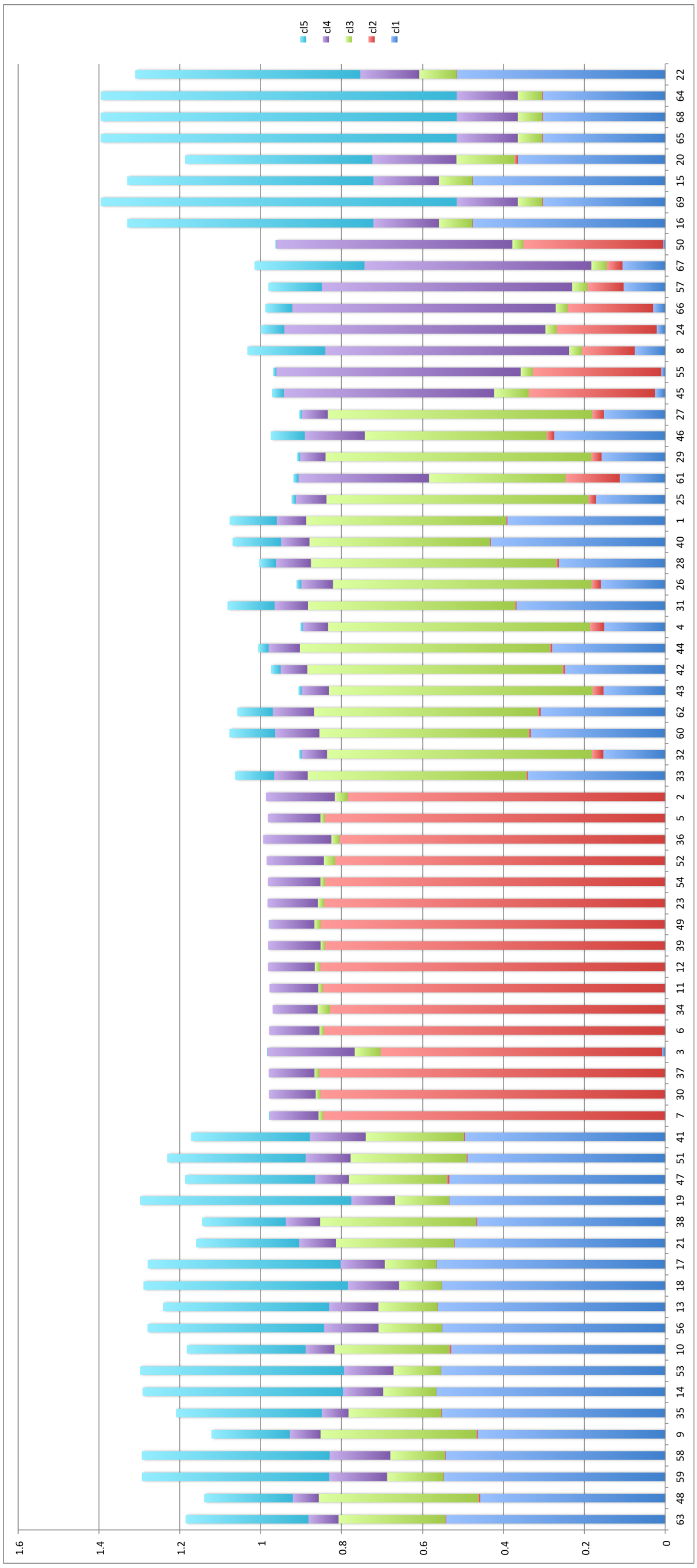


Figure 4.31: The item consensus values for each excavation site of the Guarita culture and each cluster. The higher the bar of one color the higher the consensus value for that cluster for a specific excavation site.

Figure 4.32: The same plot (Konduri) as in 4.30 but as a 100 percent plot

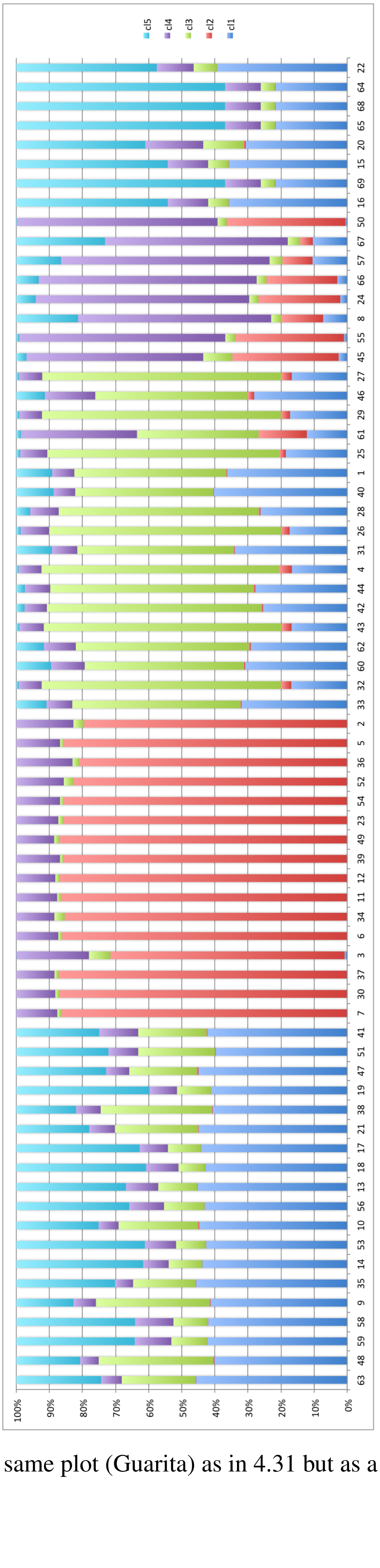


Figure 4.33: The same plot (Guarita) as in 4.31 but as a 100 percent plot

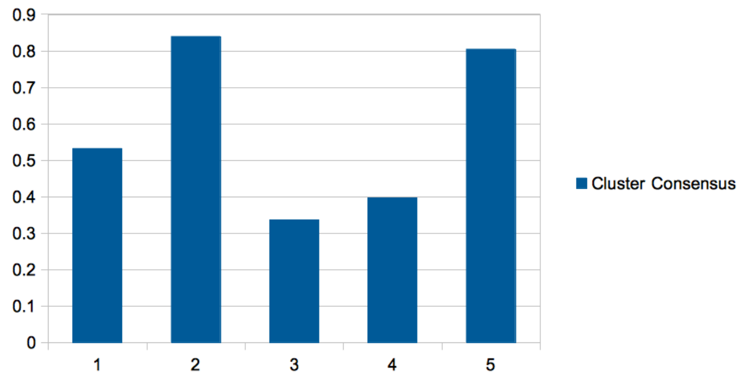


Figure 4.34: The cluster consensus values for the five clusters of the Konduri culture

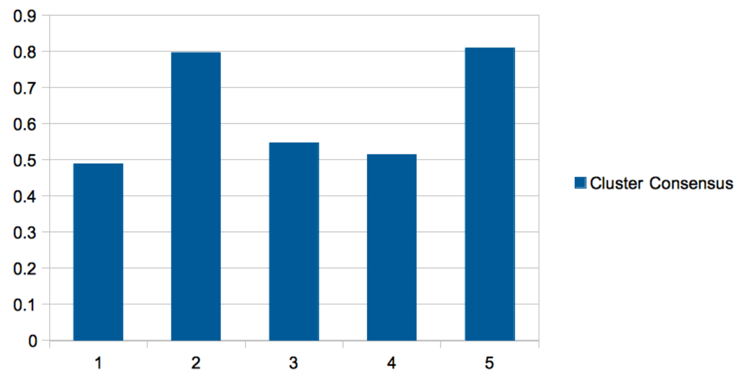


Figure 4.35: The cluster consensus values for the five clusters of the Guarita culture

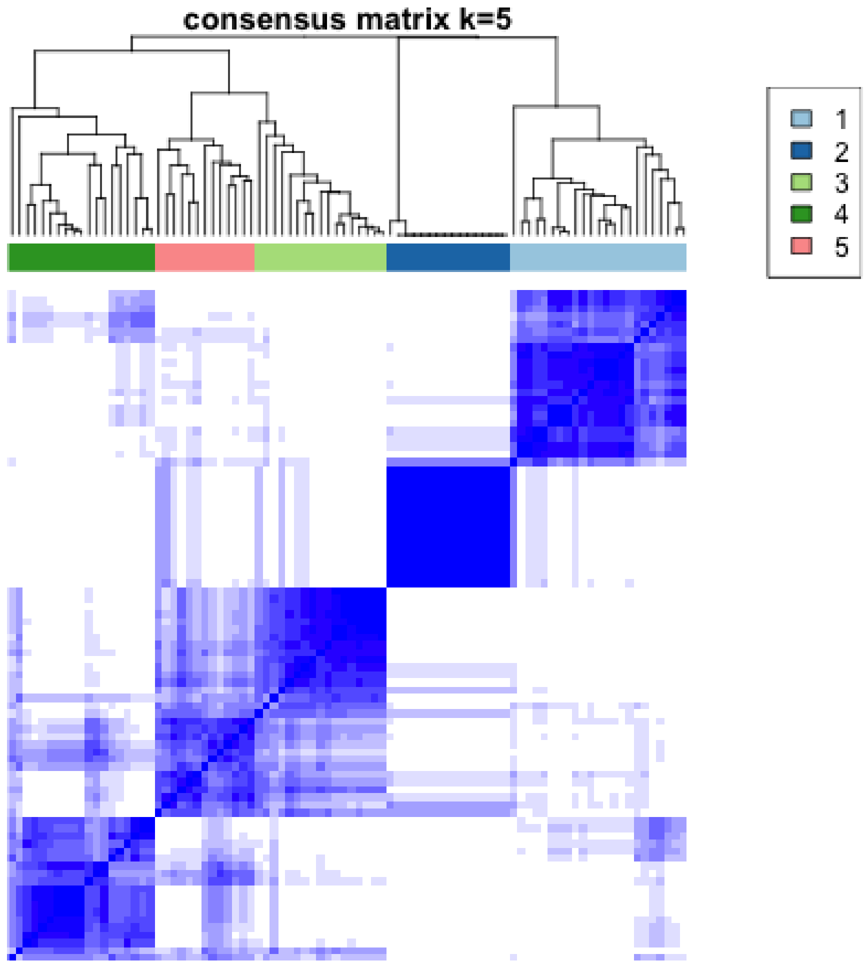


Figure 4.36: Visualization of m_{cons} combining a dendrogram and color coded values of m_{cons} for the Konduri settlements. A value closer to one (dark blue) indicates that this entry (i,j) is often grouped together in the same cluster.

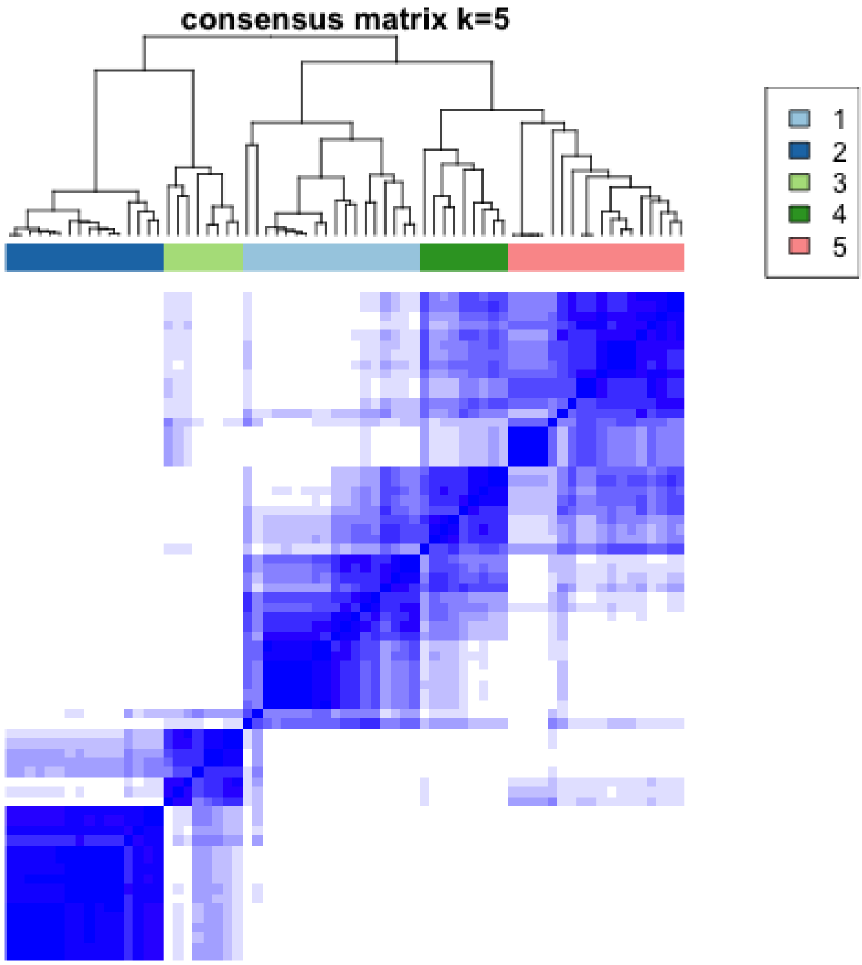


Figure 4.37: Visualization of m_{cons} combining a dendrogram and color coded values of m_{cons} for the Guarita settlements. A value closer to one (dark blue) indicates that this entry (i, j) is often grouped together in the same cluster.

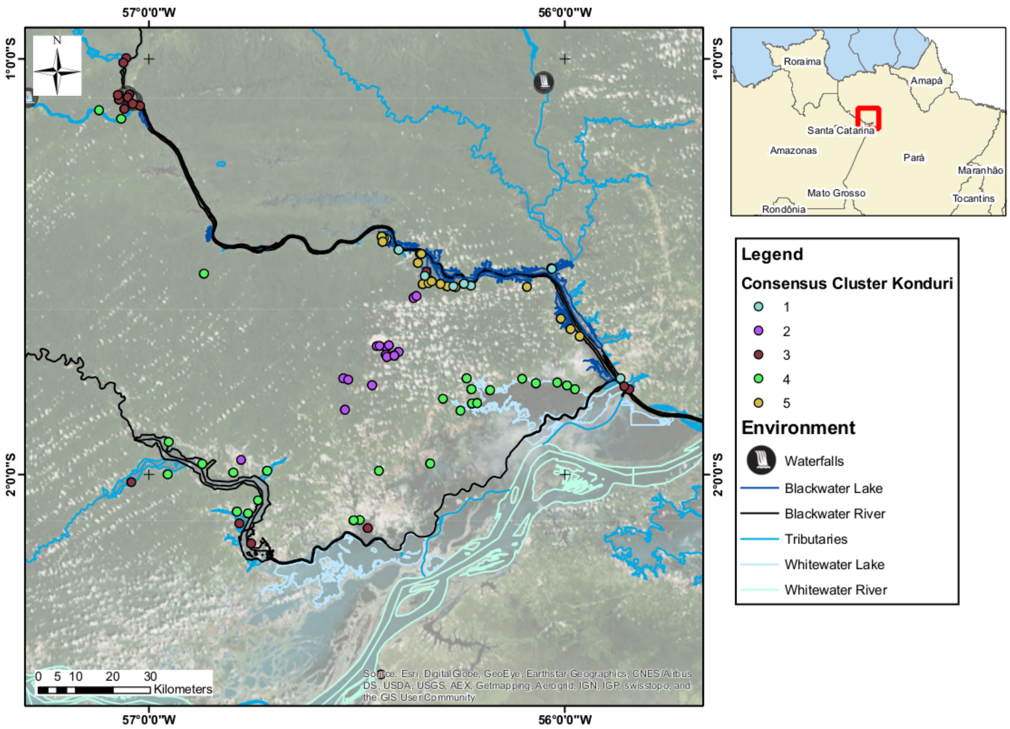


Figure 4.38: Assigning the cluster with the highest item consensus value to the excavation site for the Konduri culture

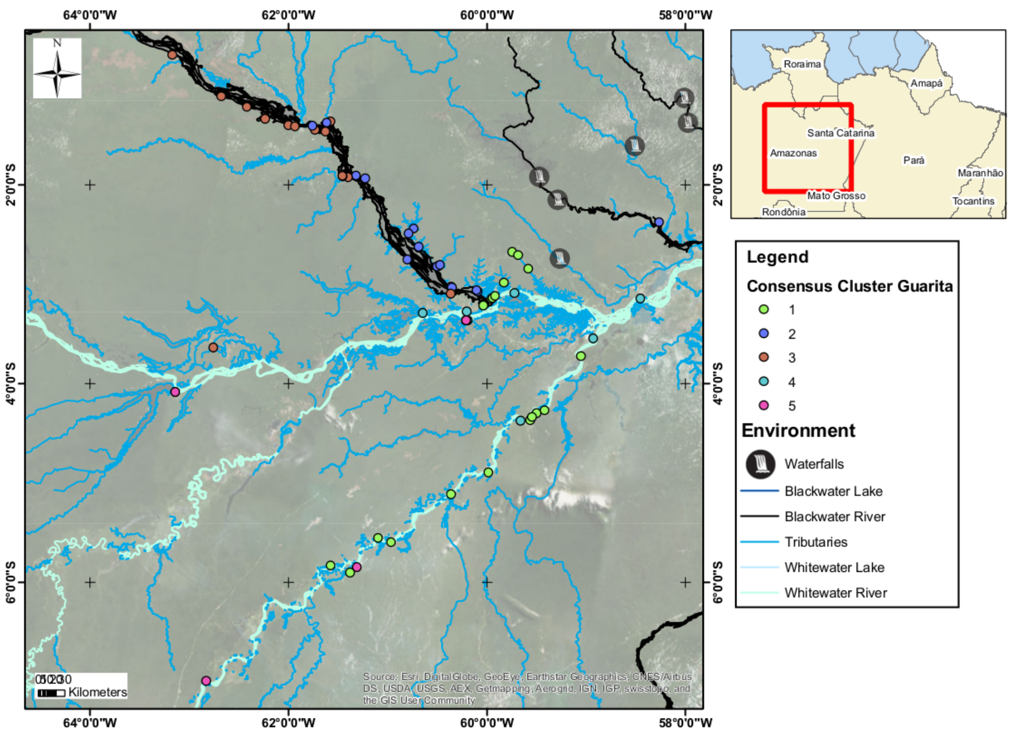


Figure 4.39: Assigning the cluster with the highest item consensus value to the excavation site for the Guarita culture

	A	B	C	D	E	F	G	H
1		max_clust	item	cl1	cl2	cl3	cl4	cl5
2	1	cl1	51	0.2893894994	0.0060555556	0.1085238095	0.1724603175	0.1533673469
3	2	cl1	46	0.2591880342	0.0033690476	0.2395555556	0.1294268078	0.1229024943
4	3	cl1	62	0.2895970696	0.0056626984	0.1064761905	0.1708024691	0.1568253968
5	4	cl1	34	0.2934920635	0.0028412698	0.145984127	0.1490299824	0.1575510204
6	5	cl1	68	0.2132661783	0.0064206349	0.1251269841	0.1841005291	0.1878344671
7	6	cl1	57	0.2613797314	0.002952381	0.0905079365	0.0215167549	0.2215646259
8	7	cl2	63	0.0001247166	0.8140893901	0.0386349206	0.0091622575	0
9	8	cl2	60	0.0001247166	0.8140893901	0.0386349206	0.0091622575	0
10	9	cl2	82	0.0001247166	0.8140893901	0.0386349206	0.0091622575	0
11	10	cl2	83	0.0001247166	0.8140893901	0.0386349206	0.0091622575	0
12	11	cl2	35	0.0001247166	0.8140893901	0.0386349206	0.0091622575	0
13	12	cl2	54	0.0001247166	0.8140893901	0.0386349206	0.0091622575	0
14	13	cl2	84	0.0001247166	0.8140893901	0.0386349206	0.0091622575	0
15	14	cl2	24	0.0001247166	0.8140893901	0.0386349206	0.0091622575	0

Table 4.1: Output of the consensus clustering. The item consensus values are plotted for each cluster and excavation site (item).

Environmental Parameters for Cluster 2 – Guarita					
	Distance to				
Normalized Item Consensus	Tributaries	Tertiary Quaternary Sedimentary Deposits	Rain Forest (Dense)	Blackwater River	Clayey
1	299.18077614	798.800871077		108.211812173	798.800871079
0.988	2937.36772035	798.800871077	98.924882426	333.83988412	798.800871079
0.987	2937.36772035	798.800871077	189.918286556	577.214929151	798.800871079
0.97	2937.36772035	1155.01414998	1155.01414999	577.214929151	1155.01414998
0.95	2937.36772035	1155.01414998	1155.01414999	577.214929151	1155.01414998
0.94	2937.36772035	1232.71360167	1232.71360167	992.115150824	1919.80088612
0.936	2937.36772035	1232.71360167	1232.71360167	2884.61206078	out of bounds
0.933	2937.36772035	1232.71360167	1232.71360167	2884.61206078	out of bounds
0.921	2937.36772035	1232.71360167	1232.71360167	4762.59571177	out of bounds
0.92	2937.36772035	1232.71360167	1232.71360167	4762.59571177	out of bounds
0.85	3847.58208107	1232.71360167	1232.71360167	4762.59571177	out of bounds
0.77	out of bounds	1232.71360167	1232.71360167	4762.59571177	out of bounds
0.71	out of bounds	1232.71360167	1232.71360167	out of bounds	out of bounds
0.59	out of bounds	1232.71360167	1232.71360167	out of bounds	out of bounds
0.1	out of bounds	1279.70701048	1279.70701047	out of bounds	out of bounds

Table 4.2: All relevant environmental properties and their distance values for cluster 2 based on the item consensus values

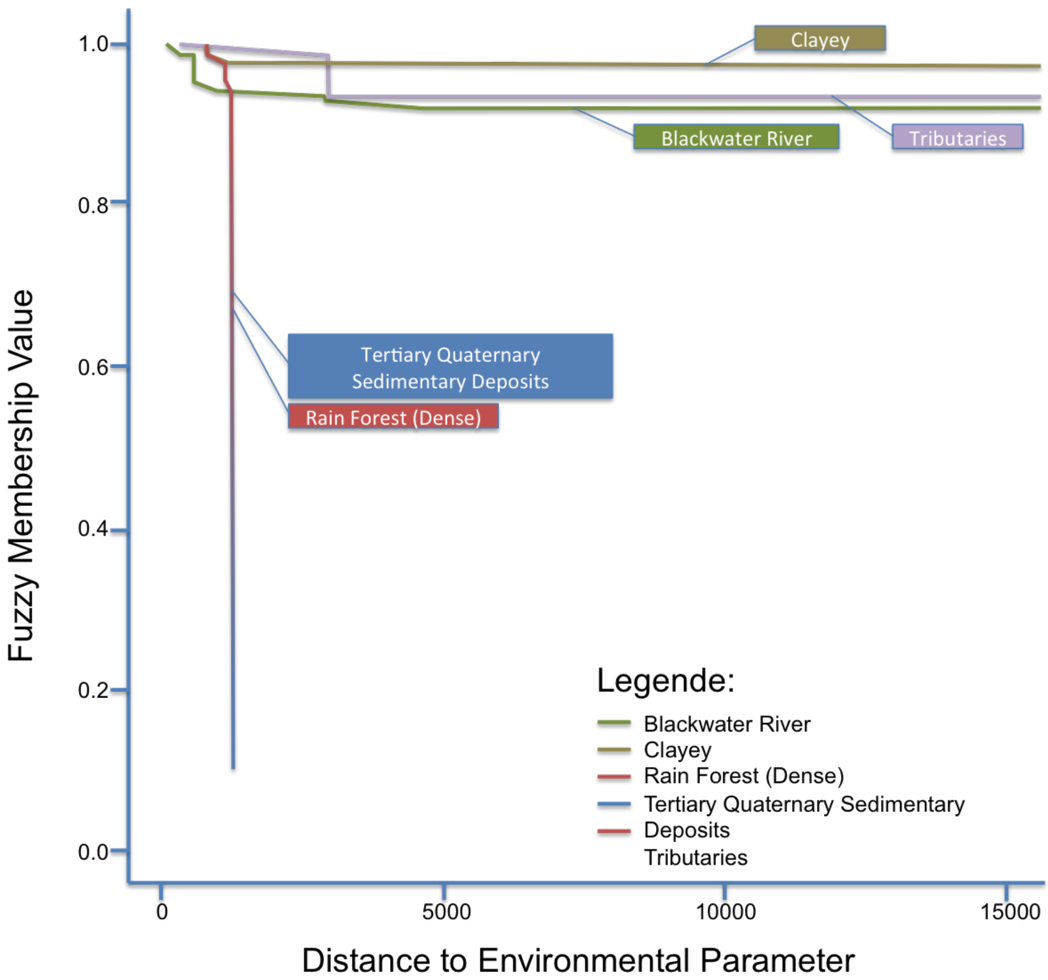


Figure 4.40: Example of a fuzzy membership function. The strength of membership is plotted on the y-axis whereas the x-axis shows the related distance of the environmental parameter.

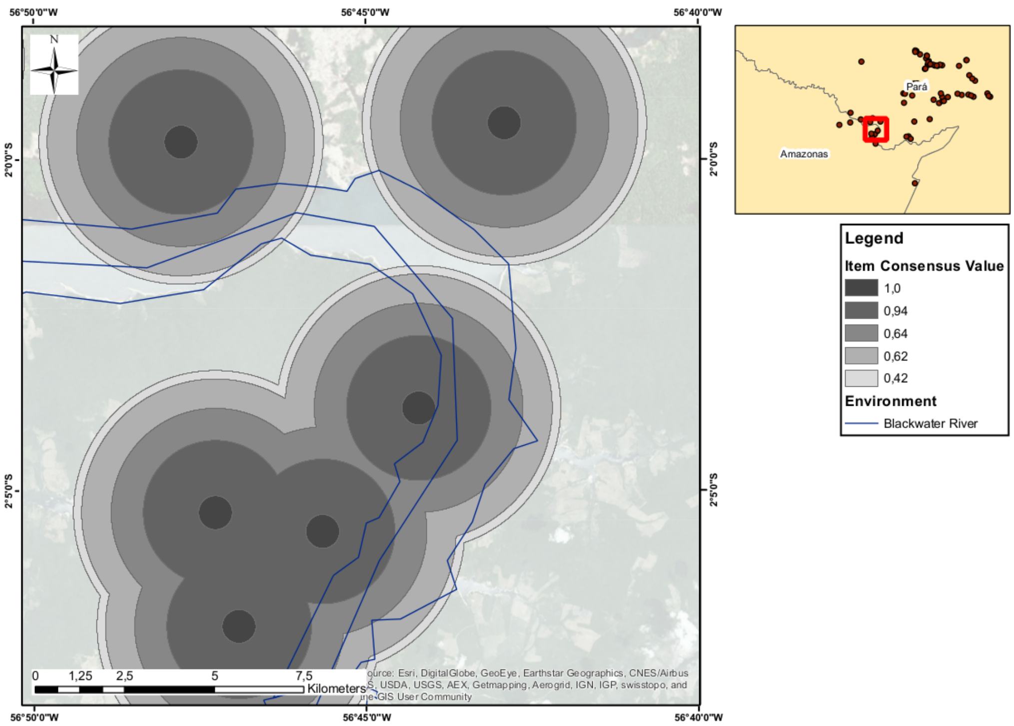


Figure 4.41: Assigning the distances to each excavation site according to the optimal cluster, environmental parameter and item consensus value. The item consensus related distances to blackwater rivers for a specific settlement type are shown.

Konduri	Cluster Number 1	2	3	4	5
Environmental Variables	Average Blackwater_Lake Blackwater_River Clayey Latossolo Neossolo Tributaries Altitude Slope	Clayey Latossolo Slope	Argissolo Average Blackwater_Lake Blackwater_River Clayey Latossolo Tributaries Waterfalls Altitude Slope	Average Blackwater_River Latossolo Slope	Blackwater_Lake Blackwater_River Clayey Latossolo Neossolo Tributaries Altitude Slope

Guarita	Cluster Number 1	2	3	4	5
Environmental Variables	Average Indiscriminate Neossolo Rain_Forest__Dense Tertiary_Quaternary_Sedimentary_Deposits Tributaries Whitewater_River Precipitation Altitude Slope	Blackwater_River Clayey Rain_Forest__Dense Tertiary_Quaternary_Sedimentary_Deposits Tributaries Precipitation Altitude Slope	Argissolo Blackwater_River Clayey Gleissolo Plintossolo Rain_Forest__Dense Tertiary_Quaternary_Sedimentary_Deposits Precipitation Altitude Slope	Argissolo Average Gleissolo Indiscriminate Neossolo Rain_Forest_Dense Tertiary_Quaternary_Sedimentary_Deposits Tributaries Whitewater_River Precipitation Altitude Slope	Argissolo Clayey Gleissolo Indiscriminate Neossolo Plintossolo Rain_Forest__Dense Tertiary_Quaternary_Sedimentary_Deposits Tributaries Whitewater_River Precipitation Slope

Table 4.3: Cluster numbers and its predominant environmental variables (item consensus value between 1 and 0.8) for the Konduri and Guarita cultures

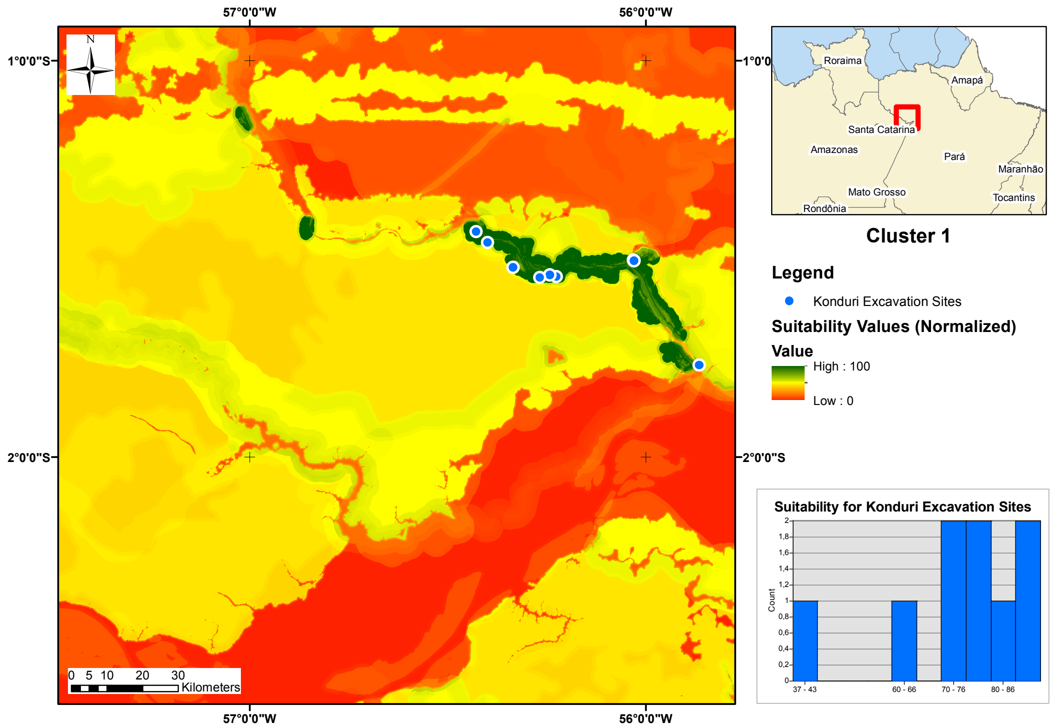


Figure 4.42: Suitability surface for cluster one. The raster values are normalized whereas a high value indicates a high suitability.

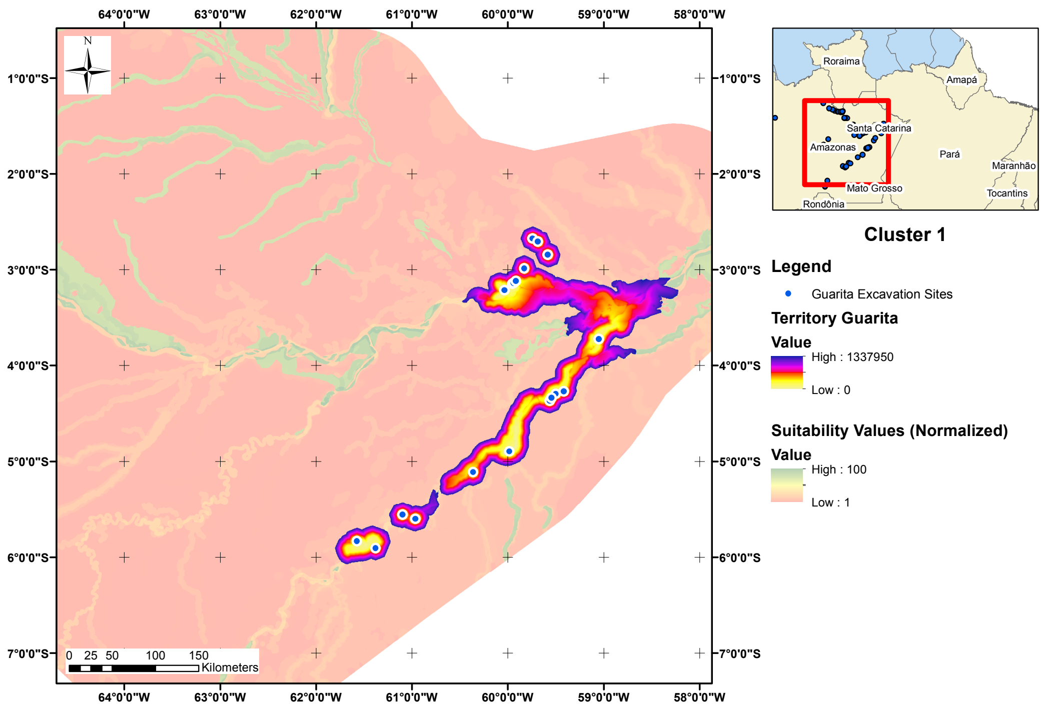


Figure 4.43: Calculated territory for cluster 1 of the Guarita culture. The raster values are based on CD_{max} whereas a high value indicates a higher cost.

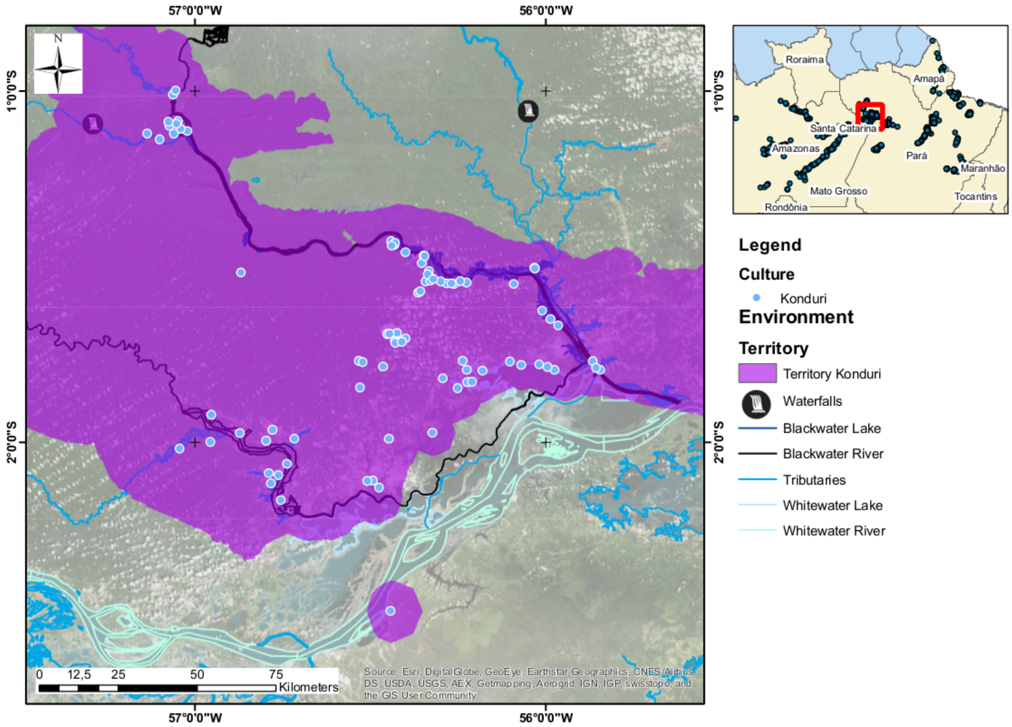


Figure 4.44: Calculated territory of the Konduri culture. The settlement function related territories are merged together to get a final territory.

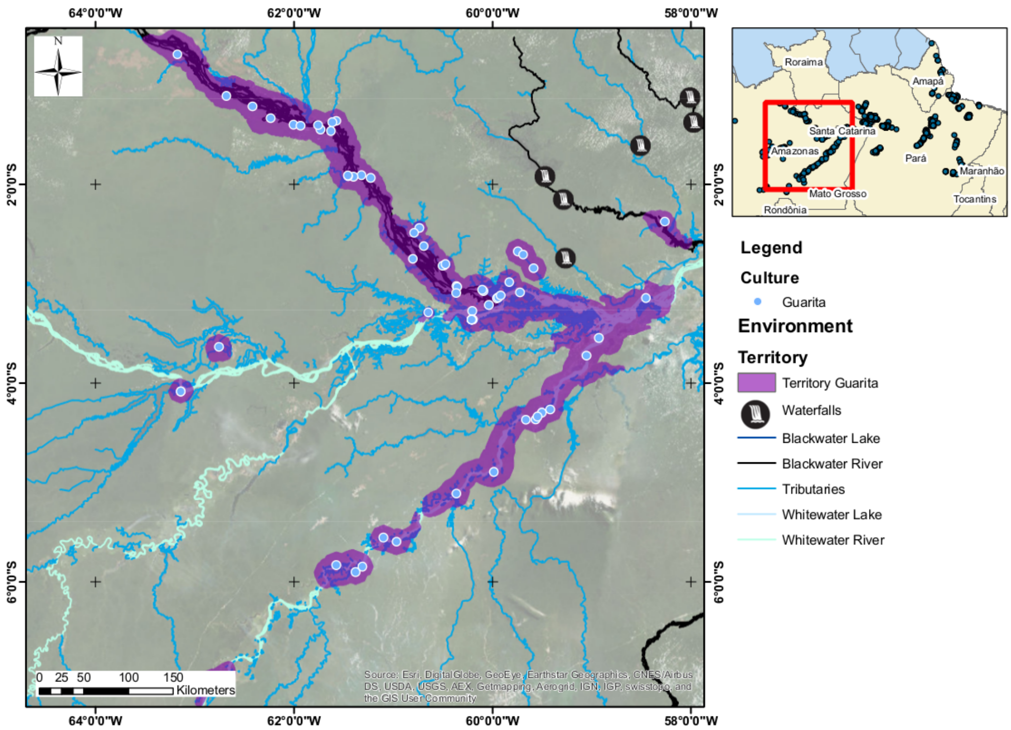


Figure 4.45: Calculated territory of the Guarita culture. The settlement function related territories are merged together to get a final territory.

5 Results and Discussion

The presented methodology is a combination of statistical methods in order to gain more knowledge about the functional settlement patterns of pre-colonial cultures in the Amazon Basin. It relies on statistical measures in order to minimize user interaction and subjectivity. The following figure 5.1 depicts the developed methodology scheme. There is still a lot of uncertainty according to the archaeological theories about functional settlement patterns. Due to that, a mainly data driven approach is developed in order to avoid expert knowledge which may lead to misleading assumptions. Nevertheless some input parameters need to be set manually. This also allows the testing of various scenarios and the identification of influencing parameters.

The data is stored in a spatially enabled database which allows to avoid proprietary data formats. The storage in a database reduces the memory due to the minimization of redundancies. Additionally the data can easily be provided on a website and can be accessed from other computers by running the database on a server instead of using local storage. Another advantage of using a database combined with server technology is the reduction of data loss in case of hardware failures. A database easily allows to mirror the tablespaces (e.g. the database tables) in order to have an identical copy of the database as a backup. Using a database implies on the one hand, that a data model is needed which is tailored to the characteristics of archaeological data and on the other hand that an import routine needs to be developed which facilitates the usage of the database. The data model is very important in order to fit the needs of database normalization, namely the reduction of data redundancies and the improvement of data integrity. Consequently, the import routine needs to convert the data (which is collected using an Excel file) so that the constraints of the data model are taken into account. The presented data model (which is described in chapter 3.1.2) is designed to fit the needs of the 3NF (third normal form).

In order to consider environmental data in the analysis, statistics as well as spatial analysis is used. The selection of environmental variables is based on literature research and archaeological expert knowledge. It is important to point out that the

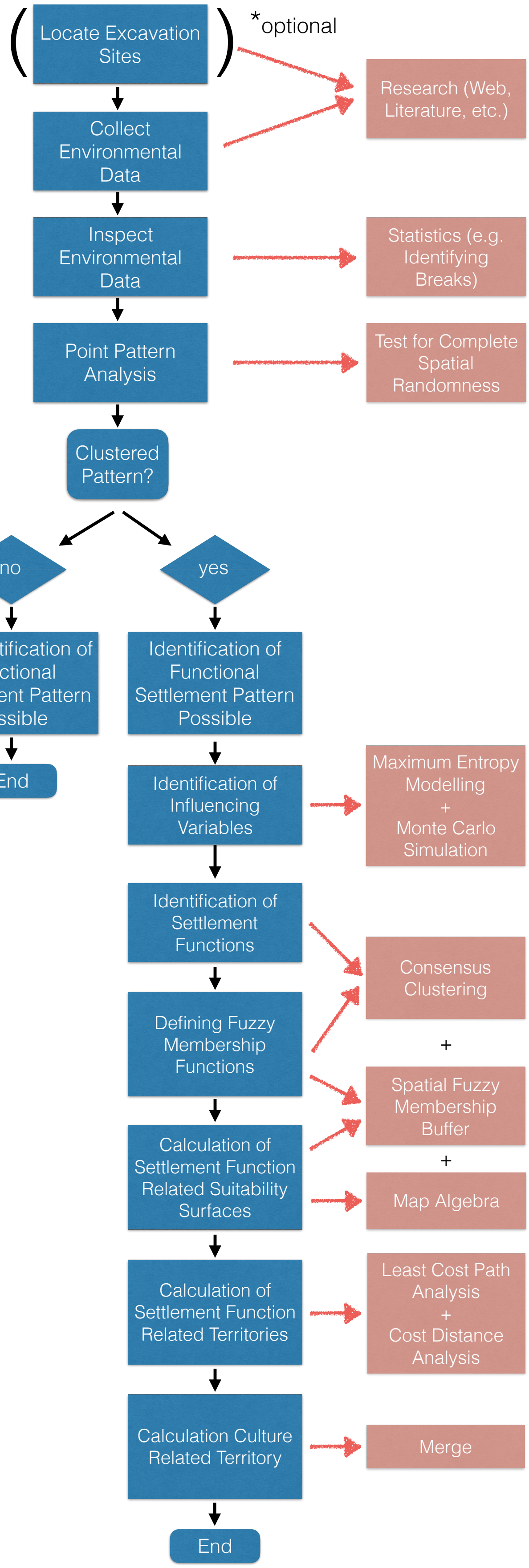


Figure 5.1: The methodology scheme which is used in order to derive functional settlement pattern related territories

chosen data is used in order to be able to test the developed approach. As described in chapter 2.2, the influencing parameters vary according to the spatial, temporal and cultural differences and thus are difficult to estimate generally. This research does not claim to use the perfect parameter set for functional settlement pattern analysis in the Amazon. However, the chosen data allows to show the potential of the developed approach. To avoid biased results due to environmental variables the presented methodology allows to reduce the number of input datasets based on statistical methods. The usage of boxplots is still an interpretative – therefore subjective – method but helps to get a first impression. In contrast, the MEM is an automated process which additionally allows to determine a variable contribution value for the environmental parameters. These values determine the influence of each variable for the resulting predictive model of the MEM. In order to avoid subjective interpretations, the variable contribution values are used as approximate importance values of the environmental variables. The environmental parameters also function as input for the consensus clustering and thus, the result is depending on them. As mentioned in chapter 3.2.3, only parameters within a specific distance are considered. This is necessary to avoid that distant variables bias the identification of settlement functions. The maximum distance is defined as a range of boundary values which allows to use a rough boundary rather than a crisp boundary. This approach allows to reduce the subjectivity and considers – to some extent – the inherently vague data but complicates the further analysis. Due to the range of distance values several cluster runs with randomly determined values – within the range of maximum distance values – need to be performed and compared in order to determine the optimal cluster solution. The results of the consensus clustering allow to draw conclusions about the quality of the clustering process. These results can be used to characterize the settlement functions based on its environmental surroundings. This is done by combining the distance measures and the item consensus values in order to define a fuzzy membership function. Thus every settlement function can be described by its distances towards environmental variables as well as its values of truth (see also chapter 3.2.4). Further analysis of the functional settlement patterns is done by calculating suitability surfaces depending on the fuzzy membership functions and the importance of the environmental variables. This leads to settlement function related cost surfaces which are the basis for the calculation of the territory.

The results of the statistical methods return precise values and serve as input for the analysis. This leads to precise boundary values or weights (e.g. distance to waterfalls for a settlement type or weights of the suitability surface), which can be misleading. Due to the inherently vague archaeological data, the output values should not be seen as exact values but rather as approximate values. However, these precise output values are used because the change of these values (e.g. due to rounding) would lead to a falsified result.

The methodology allows to perform iterative runs which changing input variables and parameters in order to develop theories on (functional) settlement patterns. However, some drawbacks exist, due to the methodology or due to the data quality. They are explained in greater detail in the following chapters.

5.1 Archaeology

The archaeological record can be biased due to several reasons. Hilbert (1968) stated that the findings were sent back to Belem by ship whereas the expedition was continued. That means that no archaeologist was on board to take care of the ceramics. Some of the bags with findings got lost or were damaged which lead to loss of information. Other problems occur because of the environmental conditions. The areas close to the water (rivers or lakes) are usually easier to access than others located in the tropical rain forest. Additionally the visibility of archaeological remains is limited in the densely forested areas which leads to insurmountable sampling bias (Barreto 1998, Meggers and Evans 1957). Due to the humid climate only certain materials can be found whereas others (e.g. feathers, wood) were not preserved.

> Due to erosion and sedimentation processes, the prehistoric evidence in mountains and sedimentary basins are often underestimated. (Pizziolo 2015, p. 13f)

Unfortunately, there is no temporal information assigned to most of the sites. This leads to the inability to distinguish whether settlements existed at the same (or an overlapping) time period or not. Due to the missing temporal information no subdivision was made. The assumption that all settlements of one culture persisted over the time of the cultural existence is probably wrong. This may cause a biased representation of subsistence strategies as well as settlement densities. Bias due to modern building activity or ploughing is another source of error for archaeological research. The so-called conventional surface (usually about 2 metres deep) is often heavily affected by human intervention (Casarotto 2015). The Amazon basin is still a scarcely populated area and only a limited number of regions are affected (e.g. the areas around the cities Manaus, Santarem or Belem). Another reason can be the sampling strategy and survey intensity of the observed settlement (Casarotto 2015).

Further errors occur due to different survey techniques. The limited number of supervising archaeologists (as shown in figure 3.6) reduces the bias due to different survey strategies.

> Unfortunately, modern models of pre-Columbian ethnic boundaries and political units have shown frequent misuse of historical data, haphazard treatment of source (including the use of secondary instead of available primary sources), and misconceptions about social, cultural, and demographic change brought by early colonial contact.(Porro 1994, p. 80)

These variations in data quality are not considered by the presented methodology. If the bias (the inaccuracies based on missing data) is known, it can be considered in the MEM by using a biased background sample. The available dataset has no information about absent or wrongly assigned findings, hence a consistent bias across the observation area is assumed. That also means that the bias due to the limited visibility and accessibility of dense forest areas is not considered. It is not known whether the available collection of excavation sites is a proper representation of the former settlement density or whether a shift towards the terra firme would be more realistic.

Nimuendaju (1952) as well a Koch-Grünberg (1923) observed that some settlements were connected by paths or roads (1 - 1,5 metres wide). More detailed information is not available, therefore the course and location of the paths cannot be considered in this approach. If – e.g. due to remote sensing techniques – the path course can be reconstructed, it can be factored in by assigning different weights to those tracks which allow better accessibility.

One aim of this research was the consideration of different functional settlement patterns. It can be assumed that the relevant environmental variables vary according to the capabilities and needs of the culture. Consequently not all excavation sites should be analysed in one step. Because of that argument, a distinction into cultures and traditions is used. Here, this is based on the variations in the ceramic pattern (see also chapter 2.1). As mentioned above, this division is only one way to subdivide the different findings into groups. Other distinctions consider the language or other behavioural aspects. Clark et al. (2005) criticised the

> site- and artifact-centered European Paleolithic studies and the lack of thinking about regional social systems (Kowalewski 2008, p. 232)

Thus, the database schema is designed so that another distinction can easily be implemented and used in the analysis. The table cultures (and similarly the table traditions) are not limited to a distinction based on ceramic styles (no underlying value table limits the possible values). The only input needed is the name of the specific group or system. This means that the input excel-file which is used to collect the archaeological record in a first step can be used. It only differs in the values which are used in the column culture. After the import of the excel file the new cultural division is stored in the database. Further calculations are based on the database records, hence no further modification is needed.

5.2 Functional Settlement Pattern

The presented approach tries to identify culture and settlement type related differences based on potentially influencing parameters. One important task is the selection of relevant variables. Whereas the methodology allows to limit the number of variables based on statistical measures, it cannot assure that all relevant influences are considered. It is possible that (yet unknown) other factors influenced the positioning. Some information may be lost such as the existence and position of sacred sites or knowledge about the necessity of a specific resource. Even if all relevant parameters are considered, no conclusions about the causality of those and the location can be drawn.

> Successful applications of data mining are not common, despite the vast literature now accumulating on the subject. The reason is that, although it is relatively straightforward to find pattern or structure in data, establishing its relevance and explaining its cause are both very difficult problems. Furthermore, much of what can be 'discovered' may well already be common knowledge to the expert. (Yuan et al. 2004, p. 367f)

Similarly, a clustered pattern towards environmental variables might not be very surprising to the archaeologist. Nevertheless, it is an important preprocessing step in order to derive reliable results. A point pattern analysis can only indicate if a pattern is random, clustered or uniform. As long as it needs to be assumed that the known location are biased no concrete proposition can be made. It is not certain whether the known locations form a representative pattern. As mentioned above, the bias may be due to the lack of visibility and accessibility of the archaeological findings, also implying an irregularly dispersed bias. Additionally no validation of the results are made which delimits the validity of the developed method.

The analysis of functional settlement pattern focusses on environmental resources which are assumed to be important for a settlement. The choice for a specific location is usually far from being random but rather characterized by the variables of the place (Barceló et al. 2015, Deravignone et al. 2015). Several theories exist about how people's decisions are determined by environmental influences whereas the environment is a dynamic system which changes permanently.

> However, for a variety of reasons, including geomorphic change, climatic change, fluctuations in sea level, and drastic changes in resource distribution with the introduction of modern land use practices, modern data may be highly unreliable. (Roper 1979b, p. 127)

This is confirmed by Van Leusen et al. (2011) who stated that prehistoric landscape is often a hidden landscape. The used environmental data is static and neither seasonal nor other changes (due to climate change, human interference, etc.) are considered. As mentioned above, the collected data includes a time span of about 6000 years and the environment is changing during such a time period. Therefore one problem is the lack of environmental data (digital as well as analogue). No reliable historic geographical data is available. However, it can easily be considered within the methodology if existent. The environmental data is stored as shapefiles (and can be uploaded to the database if wanted) and can easily be exchanged or extended. A bigger problem is the consideration of variation within the occurrence of one observation period. For now only one static environmental set of input data is used. This is probably without any problems for environmental variables such as soil type, soil condition or slope, but it definitely has an influence on the distance to water (due to flooding) or precipitation. Especially the fluvial plain várzea is an area with varying usage possibilities based on seasons. It is known for some cultures that they established complex strategies of wetland management (Heckenberger and Neves 2009). A possible solution would be to additionally consider dynamic variables. However, those variables could only be used as one static value (e.g. precipitation in the wet season, or the distance to the várzea).

> Here, as elsewhere, failure to record relevant biological information limit the conclusions which can be drawn from a statistical analysis (Diggle 2014, p. 2)

Most of the used environmental variables are used as distance measures between the settlement and the variable itself. This is not surprising due to the necessity of environmental conditions close to the site (availability of food, soil conditions or tools). In this case Euclidean distances are used. A more realistic measure might have been some sort of cost distance (e.g. travel costs), because the consideration of existing (e.g. river) networks would be possible.

Edge-effects occur, if processes are part of a larger area than the observed one. As a result, possible interactions with the unobserved areas can usually not be taken into account properly (Diggle 2014). In order to avoid edge effect of the statistical methods, an extended area is analysed. A buffer distance of 100 kilometres is used. This does not necessarily mean that the cultures never overcame this distance (e.g. for trading purposes) but it is assumed that the distant resources had no influence on the settlement location. Another aspect might be the influence of existing settlements nearby.

> Altschuh is clearly correct when he says, in the next chapter of this volume, that "magnet sites" may significantly affect settlement density in their neighborhoods, presumably for reasons that go far beyond factors of the physical and biotic environments. [...]. And yet, it is possible to find examples in the archaeological record where precisely the opposite effect has been documented. (Kohler 1988, p. 20)

The distances to neighbouring settlements is not yet considered. It might be an interesting variable to look at because the presented approach identifies variables that seem to be relevant and discards the irrelevant ones. Thus the "magnet sites" can possibly be identified in one or more settlement types.

5.2.1 Maximum Entropy

A limitation of the presented approach is the lack of absence data. The effect on sample bias is stronger, if no reliable absence information is available. Especially geographic features, such as roads or (in this case probably more relevant) river networks often lead to a sample bias (as mentioned above). A main alternative in order to limit the effects of sample bias is to use background data with similar biases or a bias grid (Elith et al. 2011). Besides the input locations a suitable set of features is important in order to develop the ecological model.

> Indeed, the constraints imposed by the features represent our ecological assumptions, as we are asserting that they represent all the environmental factors that constrain the geographical distribution of the species.(Phillips et al. 2006, p. 237)

The results of the MEM only indicate potentially influencing variables but cannot determine the causal relationship between those variables and the settlement location. Other statistical methods may lead to different results and for now it is not possible to test which of the statistical models returns the most reliable results. A

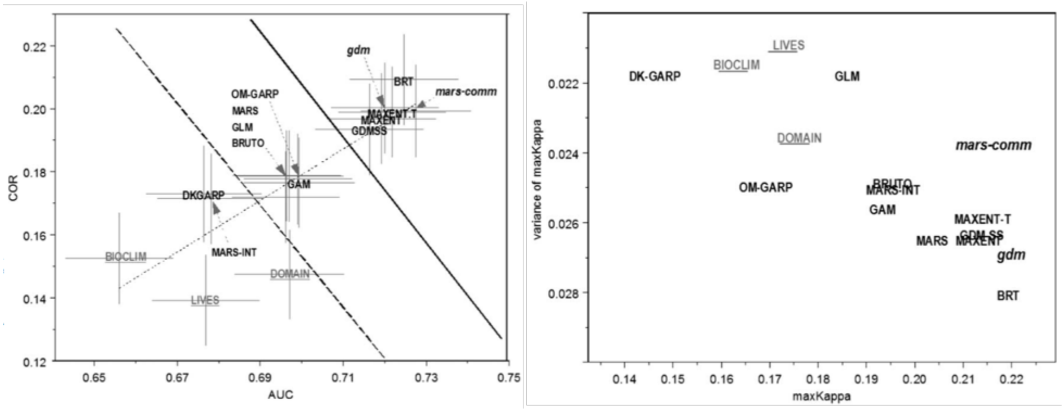


Figure 5.2: The maximum entropy model compared to other predictive modelling techniques (Elith et al. 2006)

comparison of several predictive modelling techniques and case studies show that maximum entropy modelling performs well when averaged across species and regions (Elith et al. 2006). As shown in figure 5.2 maximum entropy modelling ranges among the highest-performing methods for most species. The figure on the left plots the mean AUC versus the correlation. The correlation is similar to the AUC but additionally carries information about how far the prediction varies from the observation (Murphy and Winkler 1992). The figure on the right uses the chance-corrected measure kappa which is commonly used in ecological predictive modelling. The index considers omission as well as commission errors (Cohen 1960). The models perform best if the variance of maximum kappa is low whereas the value itself is high (top right of right figure 5.2). The results compare the model accuracy according to predictive modelling, which is not the considered in the used methodology in this study. Nevertheless, a high performance of the model is necessary in order to use the results as input values for further analysis. As shown in chapter 4.4.1 both case studies have a high performance, whereas the Konduri model has a higher performance.

5.2.2 Clustering

Only the variables with a permutation importance value > 0 are used in the clustering analysis. The clustering is used in order to determine similar groups, namely different settlement types. The result is the identification of groups which are similar (based on a similarity measure) based on the given input data. The reliability of cluster results depends on the quality of the input data. It is assumed that the settlement type is based on geographical characteristics which meet the requirements of the settlement function. Due to that, a combination of statistics and literature review (which is the consideration of expert knowledge) is used in order to derive a decisive set of input variables. It is difficult to consider social or religious influences (such as ritual habits or the distance to sacred places) due to two reasons. The first reason is the lack of knowledge for most of the observed groups. The other reason is more related to the presented approach. It can be assumed that those influences vary according to the observed culture. Thus an expert needs to identify potential influences beforehand. This means, if those habits can be described in discrete or continuous values (e.g. distance to next settlement, type of ritual habit, spoken language, etc.) they can be considered in the analysis. This allows the expert to test various scenarios and see which variables have influence on the functional settlement pattern.

In this case study a partitioning clustering algorithm, namely k-means, is used in order to determine the settlement functions. This is a commonly used algorithm with some drawbacks. The result of the clustering process does not necessarily return the best result but it is a very fast algorithm. Due to the consensus clustering approach and the associated high number of cluster runs, the processing time is a relevant aspect. The result of a single k-means cluster run depends on the selection of the starting points. In order to avoid misleading results, the clustering process should be repeated several times using varying starting points, which is done in the consensus clustering. Another influencing parameter is the number of clusters k. Choosing an unsuitable value for k may lead to an unintuitive solution. The consensus clustering was initially designed for the comparison of clustering results using a different number of clusters. Although this research adapts the consensus clustering in order to identify the most likely clustering result, it can also be used for the determination of the optimal cluster number. The consensus clustering, when used to determine the optimal number of classes, can lead to over-interpretation of cluster stability (Şenbabaoğlu et al. 2014). In order to avoid that, the decision about the optimal number of clusters can be determined using other methods (e.g. optimum average silhouette width).

Other disadvantages are the disability to identify outliers and hierarchical clusters (clusters within a cluster). Whereas the latter aspect is not relevant in the case of settlement types of prehistoric cultures, the former can lead to misleading results. In order to avoid this, a noise reduction can be applied beforehand, or another algorithm needs to be used. The consensus clustering function which was adapted to the specific needs of the settlement type and nearness detection is able to use various clustering algorithms.

The results rely on the chosen maximum and minimum boundary values $max_{near_{outer}}$

and $max_{near_{inner}}$. Those are determined based on previous archaeological research but vary according to the spatial, temporal and cultural properties. Thus a change of those values can lead to a new assignment of settlement types.

5.2.3 Suitability Surfaces and Territories

The result of the least cost path relies on the underlying cost surface. Thus the calculation of the cost surface is an important task. The suitability values of cluster two of the Konduri culture are either very high or very low. This is caused by a high difference of the used parameter weight. This leads to predominant environmental variables whereas others are not influencing the suitability values even though they were selected as influencing parameters in the consensus clustering process. A similar effect can be seen in cluster 4 of the Konduri culture. In this example a few excavation sites are located in unsuitable areas due to the high weight of the other environmental parameters. Thus the least cost path lead to very high maximum cost value CD_{max}.

The cost surfaces are also the basis for the territorial analysis. This means that wrongly assigned suitability values lead to errors in the resulting territory. The calculation of the cost surface is based on the results of the MEM as well as the consensus clustering. Thus inaccuracies in one or both methods lead to a misleading output in the end.

The cost surface shows whether a location in the study area provides suitable environmental properties or not. It does not provide information about the willingness and ability to overcome unsuitable areas in order to have access to suitable areas. The environmental variables are another aspect which may distort the analysis. As mentioned above, some cultures had the abilities to use the wetlands for agricultural purposes. This means that for some cultures the várzea can be – to some extent – suitable for farming. Due to the fact that only one settlement of the observed Konduri culture is close to this fluvial plain, the várzea is not considered to be suitable in that case. Due to the small sample size of one excavation site no meaningful statement about the suitability can be made. Only assumptions about the influencing and preferred variables can be made. Thus the method facilitates archaeological research and allows to develop new scenarios or supports existing settlement theories. However, expert knowledge is still required, not only to set some input parameters but also for the interpretation of the results.

The territories are sometimes not connected. That does not mean, that no relationships with other settlements occurred but rather that their territories (the exploited areas) are not connected. Another reason can be the lack of data. Only one former Konduri settlement is located south of the Amazon river. Because of the low suitability of the river and the várzea (due to the floodings), the territory seems to be isolated from the others. A similar situation can be identified with the settlements further away from the rivers. Due to the lack of a supposedly suitable environment, only small isolated territories were calculated. In this case the used input dataset is important. Those settlements are all located next to (sometimes small) rivers which are not represented in the river network used for this study. This means that the resulting clusters rely on the quality and integrity of the input data. As can be seen in figure 4.44 especially the results for Konduri seem to cover almost the whole extent of the study area. This is mainly caused by the suitability surfaces of cluster two and cluster four. Three main reasons can lead to an overestimation of the territory:

1. At least one excavation site is located in an unsuitable area. This leads to a very high CD_{max} value.

2. A high weight is assigned to one environmental parameter. Thus the presence of this specific environmental parameter determines whether an area is suitable or not.

3. Only a few item consensus values (e.g. only 1.0 and 0.1) are influencing the calculation of the suitability surface. The abscence of medium values leads to an either high or low suitability.

6 Conclusion

6.1 Summary

The aim of the research is to develop a methodology which allows to gain information about functional settlement pattern of precolonial cultures. The approach can be subdivided into four major subsections, namely the location allocation and storage of the archaeological record, the identification of environmental parameters and their influence, the determination of functional settlement patterns and the calculation of the settlement function related territories.

The locations of the former settlement serve as basis for the approach – which is designed to be mainly data driven in order to avoid subjective interpretations. Therefore a first goal was to establish a data basis with reliable location information of the former settlement. The settlement locations are derived from research articles where either a coordinate or – which is mostly the case – a description and/or a map of the location is given. This means, that only a point is (more or less exactly) known after the location allocation process. The coordinates were mainly determined by expert knowledge (namely Klaus Hilbert who is an expert in Amazonian Archaeology). The manual processing is necessary due to the lack of reliable digital data sources. The database which is provided by IPHAN does not have any coordinates and can therefore not be used. The database presented by WinklerPrins and Aldrich (2010) provides coordinates but neither the unique excavation site name nor information about the former cultures or traditions. A database is used to store the data. Consequently a data model was developed which is tailored to the needs of the manually recorded and located data. The database is running on a server which allows easy access from other computers (and therefore also websites) if the permission is given. The server technology is an optimal solution if sharing and providing data to others is an important goal and in this case is preferable to client based storage concepts. Another aspect is the reduction of data redundancies, avoidance of proprietary data formats and the improvement of data integrity. Due to that, a database is used which – if database normalization is applied – reduces the redundancies. The decision for a PostgreSQL database was made, because it is an open source project which stores the geometries in an OGC (Open Geospatial Consortium) compliant format. In order to collect as many former settlement locations as possible – which can be accessed and used for the analysis – only the location (as a point coordinate) is marked as mandatory field and additional information is optional.

It cannot be determined whether a causal relationship between the surrounding environmental and the settlement location exists. Due to the fact that some environmental conditions are necessary in order to survive (e.g.enough water and food) – especially in an archaeological context – it can be assumed that the environment is an important factor in order to determine how suitable a specific location might be. If assumed that humans tend to optimize the suitability, the environmental variables at existing locations allow to draw conclusions about the needs of the people who settle(d) in that region. The identification of environmental variables is based on archaeological research and is used as the most likely set of parameters. This approach allows to perform several iterative model runs with a varying set of input parameters, which is important to function as a knowledge discovery process. This allows to test for several scenarios in order to validate a specific theory or develop new settlement theories. This implies a changing set of input locations (e.g. to check whether ceramic styles or other aspects such as language have influence on the result) as well as changing environmental variables (e.g. to test for other environmental influences). The focus is on the settlement locations of a group rather than inter-site relationships, in other words a macro scale approach. The Maximum Entropy Model is used in order to determine the importance of each environmental parameter. It is a commonly used approach in terms of predictive modelling of species environmental requirements and geographic distributions. The variable contribution values for the predictive model are an important outcome of the Maximum Entropy Model. Other techniques in order to determine the influence of variables would be statistical methods such as the (multiple) linear regression or a Principal Component Analysis. A disadvantage of using a linear regression is that only linear dependencies are recognised which cannot be guaranteed. Another problem may occur due to outliers which have impact on the result. A Principal Component Analysis reduces the number of input variables by merging variables to principle components in order to get the best approximate of the input dataset. The idea of the presented methodology is to derive information about the functional settlement patterns based on the environment which is assumed to be an important resource and influencing factor. The disadvantage of the Principal Component Analysis is that further analysis is based on the principle components and not on the environmental variables itself. This complicates statements about the influence of specific variables.

> The traditional tools used by archaeologists include, respectively, linear or logistic regression and nearest neighbour or quadrat analysis, but each of these raises methodological problems. The first two have the capacity to mislead in contexts where spatial dependence can be shown to exist (i.e., most geographic contexts: Fotheringham et al., 2002:162 – 166), and the last two are insufficient for detecting multiscalar spatial patterns. (Bevan and Conolly 2006, p. 218)

It is assumed that settlements of one culture fulfilled specific functions (such as trading, defence or agricultural purposes) which can be distinguished based on the environmental conditions surrounding the site. The methodology is designed so that settlement function related results can be provided. Most of the environmental variables are determined by using their distance to the excavation sites, e.g. distance towards nearest blackwater river. Temperature and precipitation are provided as categorized variables by the IBGE, thus the categories are used as input. The various continuous and categorized values (e.g. distance measures or classes of precipitation ranges) and the observed excavation sites serve as input. In order to identify settlement types, a cluster analysis is performed. Due to the big extend of the area, the cluster result was influenced by variables in greater distances which can be misleading. A river which is more than 100 kilometres apart was probably not decisive for the determination of the settlement location. A boundary value is set which defines nearness. An environmental parameter within the given boundary value serves as potentially important resource for the specific settlement strategy. In order to consider dynamic nearness values, the boundary value is individually set for each environmental variable and culture. To stick with the data driven approach a rough boundary is defined by a range of minimum and maximum values. The cluster analysis is performed many times (1000 in this case) whereby the nearness values for each environmental variable are randomly set for each cluster run (based on the predefined boundary range). A data driven definition of resource related nearness can be made by comparing the clustering results. The results of the various cluster runs are compared using a consensus clustering approach. This approach counts how often each excavation sites pair is within the same cluster and derives an item consensus value which reflects the goodness of the cluster assignment for each excavation site. The consensus cluster value describes the goodness of a cluster itself (which basically is the mean of item consensus values for all excavation sites which are assigned to this cluster). The item consensus value is used to define the fuzzy membership function and thus the degree of truth for the resource related nearness.

Suitability surfaces for each cluster are calculated based on the fuzzy membership function. The results are suitability surfaces which are based on the assumed importance of the surrounding resources considering the settlement functions. Those suitability surfaces can then be used to calculate the settlement function related territories. This – in a last step – is achieved by calculating the least cost path as well as the accumulated suitability value which is used as a maximum value in a cost distance calculation. In contrast to other territory analysis approaches, the presented approach allows to consider the settlement functions. The MEM also returns a suitability surface, which can be used as a basis for the territorial analysis, but calculates the suitability based on all observed excavation sites. In contrast to that, the presented approach calculates several suitability surfaces – one for each cluster. This leads to settlement function related territories which are calculated only on the basis of the settlement function related suitability surface. Other approaches use geometrical techniques such as Thiessen polygons (e.g. the XTENT model). This implies that the whole observed area is subdivided into territories and geographical boundaries are difficult to consider.

6.2 Major Findings

An important outcome of this thesis is the database which is located on a server and provides access to all the allocated locations which were found in previous publications. This database is far from being complete but – for now – is the biggest collection of excavation sites which has coordinates as well as the official unique excavation site id (which is commonly used among archaeologists in Brazil). An export routine was developed which allows to export the database content into a shapefile. Consequently, the data can both be easily exchanged and be used for further spatial analysis in a GIS. Due to the usage of the official unique excavation site id this database can – in contrast to the one published by WinklerPrins and Aldrich (2010) – easily be enriched with additional information from data sources which use the id (e.g. IPHAN). Since the database is located on a server it can be configured so that it is accessible via the internet. This allows access to the data and, as a result, facilitates further research.

In case of archaeological data it is difficult to determine the settlement related influencing parameters, because it is difficult to validate the result. The ability to approximate the influence based on the variable contribution values is an objective statistical way to derive these settlement related parameters. The MEM, which is used to determine the variable contribution values, derives those values based on the locations, thus no other dependent variables are needed. The comparison of these values can be done by applying various scenarios with a varying set of potentially influencing variables.

One aim of the methodology was to gain knowledge about the settlement patterns. In order to analyse the former settlements and their environmental surroundings, a Monte Carlo simulation can be applied. This allows to test the locations against all potentially influencing variables individually. By creating randomly dispersed points within the same spatial extend a comparison of the results with the ones from the excavation sites is possible and patterns can be identified which occur only for some excavation sites. The use of a Monte Carlo simulation as well as the consensus clustering ensures statistically reliable results. The high number of model runs within a Monte Carlo simulation allows to draw conclusions about the reliability of the point pattern process. Because the Monte Carlo simulation is difficult to interpret in terms of spatial relations a clustering approach is performed additionally in order to identify settlements with similar environmental surrounding. The variation of environmental conditions (e.g. presence or absence of stones) helps to identify the settlement types. As a result, each settlement function is described by a combination of environmental variables which serve as a settlement function specific description. Additionally, the derived cluster can be assigned to each excavation site and allow to see the spatial distribution of the various settlement functions.

Based on the gained information suitability surfaces can be calculated which consider the specific set of environmental conditions relevant for the settlement function. In contrast to other approaches not one single suitability surface which serves as basis for further analysis is calculated but one for each settlement function to consider the settlement function related characteristics. The variable contribution values of the MEM are used as weights for the calculation. If suitability optimization can be assumed for cultures, this surface allows the identification of potential other settlements. These suitability surfaces serve as a basis for the concluding calculation of the territory. In order to consider previously gained insights, a new method was developed, which uses the results of the clustering (namely the maximum distance measures of each settlement function) in combination with the suitability surface. The result is a cumulative distance value which serves as a maximum cost distance value for the territory analysis.

The functional settlement pattern analysis is based on the results of the cluster analysis. Therefore, a test for complete spatial randomness needs to be done in advance. In this case a combination of R-, F- and K-functions is used. In order to get satisfying results for the cluster analysis, it was necessary to delimit the maximum distance towards the environmental variables. Neither a PCA nor a re-scaling function provided reasonable results. In order to minimize subjectivity, not a crisp value but rather a rough boundary is used. A high number of cluster runs with varying, randomly selected maximum distances (within the range of the rough boundary) were applied in order to get a statistically reliable result. A consensus clustering approach is used for the comparison of the cluster runs. As a result, a measure of robustness, thus the goodness of fit for each cluster as well as for each excavation site is returned. Those robustness measures can then be used for the fuzzy membership functions, i.e. that the uncertainty can be considered in the presented approach. This fuzzy membership function can additionally be applied to the calculation of the suitability surface and consequently has influence on the territory analysis.

Only very little user interaction is required. On the one hand environmental variables need to be chosen (such as water source, soil type etc.), on the other hand the definition of the boundary range values, namely the minimum and maximum assumed distance which a culture is willing to overcome in order to exploit a specific resource need to be set. Default values are set for each variable but if those need to be adjusted, expert knowledge, and therefore user interaction is required. This means that a default set of parameters is used to allow a complete automated data driven approach, but expert knowledge might be useful in order to gain more reliable results or test for various settlement scenarios.

6.3 Future Work

The presented approach is data driven which means the data quality of the input data is relevant. A representative sample of settlement locations is required in order to calculate reliable results. This means that the validation of the input data as well as the results is necessary. Therefore the location accuracy needs to be checked as well as the existence of yet unknown further settlements. Additionally, the collection of reliable absence data might improve the output, due to the assumed sample bias. This would lead to a dynamically improving model which adapts the presence and absence information and thus an iterative knowledge discovery process.

Another aspect which is related to the data quality is the considered set of environmental variables. In future analysis additional or other variables should be considered, which includes non-environmental information (e.g. distance to other settlements or sacred places). The water, respectively river network is far from being complete. The whole area is crossed by small channels or rivers (some of which are only active during the rainy season) and water seems to be an important resource. This is not only the case with water resources. The categorisation and resolution of all the environmental variables influence the result. Resolution in this case means not only the spatial resolution (e.g. the pixel size) but also the temporal resolution.

Figure 6.1: Interpretation of the settlement functionalities based on the cluster analysis. This functionality type describes settlements of clusters which are mainly located close to blackwater rivers. The colors indicate the category. Blue means it's a economic function, green is for social functionalities and red for food related functionalities.

Maybe the settlement location is influenced by seasonal changes which can only be identified if such variables are are considered. Additional temporal information for the excavation sites allows the analysis of movement behaviour and the identification of simultaneously settled sites. Thus, the usage of more detailed data (which also includes more computing time due to the size of Amazon basin) can lead to other – and probably more reliable – results.

In order to validate the presented approach, it should be applied to other datasets either of well researched archaeological cultures or recent cultures. The methodology can easily be adapted for other datasets. The only things that probably need to be adjusted are the (environmental) input variables as well as the assumed rough boundary values for each variable. For now, some methodologies require some data preparation steps in order to work properly (e.g. the MaxEnt software only works with CSV- and ASC-files). The needed conversion tools are developed in Python but smaller modifications are needed (e.g. other in- and output paths) in order to work properly with other datasets.

In cooperation with archaeologists, the various settlement types can be interpreted in order to define their specific functionalities. A prototype of a possible interpretation of the Konduri settlement type was made by Prof. Klaus Hilbert in 2014 (unpublished) as exemplarily shown in figure 6.1. Therefore four categories, namely economy, social, food and religion were defined and all the various functionalities are assigned to those groups. For communication as well as for further research purposes a visualization and communication platform would be helpful. It allows to provide a collection of publications and thus a kind of Amazonian archaeology related knowledge database. Additionally, the locations and other datasets (e.g. the environmental variables) can be provided. A webgis can be implemented in order to facilitate further research. This is done in a beta version which is online already (link: `http://terrapreta.geo.uni.augsburg.de`). Some exploratory tools are implemented, such as the comparison of distances towards environmental parameters for the selected sites. For now only the water resources such as waterfalls, tributaries, whitewater rivers or lakes and blackwater rivers or lakes are considered. Further tools and environmental datasets would help to build a helpful exploratory analysis platform for functional settlement pattern analysis (von Groote-Bidlingmaier et al. 2015).

For now no comparison of different functional settlement pattern analysis is implemented. Thus, only a visual interpretation of various settlement types, territories and suitability surfaces is possible whereby the comparison of territories is difficult. This is due to the lack of crisp boundaries as well as the locational differences. A possible approach would be to consider the underlying suitability surface but this is still work in progress. It would be interesting to be able to compute different functional settlement pattern based on the calculated output.

Bibliography

Abdi, H. (2003). Neural networks. In Lewis-Beck, M., Bryman, A., and Futing, T., editors, *Encyclopedia of Social Sciences Research Methods*. Thousand Oaks.

Abdi, H., Valentin, D., and Edelman, B. (1999). *Neural Networks*, volume 124 of *Quantitative Applications in the Social Sciences*. Sage Publications, Inc.

Ahmad, E. (1952). Rural settlement types in the uttar pradesh (united provinces of agra and oudh). *Annals of the Association of American Geographers*, 42(3):223–246.

Aldenderfer, M. (1998). Quantitative methods in archaeology: A review of recent trends and developments. *Journal of Archaeological Research*, 6(2):91–120.

Almeida, F. O. d. and Neves, E. G. (2012). The polychrome tradition at the upper madeira river. In Rostain, S., editor, *Antes de Orellana - Actas del 3er Encuentro Internacionalde Arqueología Amazónica*.

Andel, T. H. v., Runnels, C. N., and Pope, K. O. (1986). Five thousands years of land use and abuse in the southern argolid, greece. *Hesperia: The Journal of the American School of Classical Studies at Athens*, 55(1):103–128.

Andrienko, G. L. and Andrienko, N. V. (1999). Data mining with c4.5 and interactive cartographic visualization. In *User Interfaces to Data Intensive Systems, 1999. Proceedings*, pages 162–165, Los Alamitos.

Andrieu, C., de Freitas, N., Doucet, A., and Jordan, M. (2003). An introduction to mcmc for machine learning. *Machine Learning*, 50(1-2):5–43.

Angell, I. O. and Moore, R. E. M. (1984). On archaeological applications of the voronoi tessellation. In *Computer Applications in Archaeology 1984 (Birmingham 1984)*, pages 9–14.

Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., and Parker, M. (2002). Population growth and collapse in a multiagent model of the kayenta anasazi in long house valley. *Proceedings of the National Academy of Science*, 99:7275–7279.

Baddeley, A. (2010). Analysing spatial point patterns in r. CSIRO and University of Western Australia. Workshop Notes.

Baddeley, A. J. (1998). *Spatial sampling and censoring*, volume 2, chapter 2, pages 37–78. Chapman and Hall.

Bahrenberg, G., Giese, E., and Nipper, J. (2003). *Statistische Methoden in der Geographie. Band 2. Multivariate Statistik.* Studienbücher der Geographie. Berlin, Stuttgart: Borntraeger, Stuttgart, 2 edition.

Bailey, T. and Gatrell, A. (1995). *Interactive spatial data analysis*. Longman Scientific & Technical.

Balée, W. (1993). Indigenous transformation of amazonian forests : An example from maranhão, brazil. *L'Homme*, 33(126):231–254.

Balée, W. (1994). *Footprints of the Forest: Ka'apor Ethnobotany-The Historical Ecology of Plant Utilization by an Amazonian People*. Columbia University Press.

Balée, W. (1998). Historical ecology: premises and postulates. In *Advances in historical ecology*, pages 13–29. Columbia University Press.

Balée, W. and Erickson, C. (2006). *Time and Complexity in Historical Ecology: Studies in the Neotropical Lowlands*. The historical ecology series. Columbia University Press.

Barceló, J. A., Achino, K. F., Bogdanovic, I., Capuzzo, G., Del Castillo, F., de Almeida, V. M., and Negre, J. (2015). Measuring, counting and explaining: An introduction to mathematics in archaeology. In *Mathematics and Archaeology*, pages 3–64. CRC Press.

Barnard, G. A. (1963). Discussionon professor bartlett's paper.

Barreto, C. (1998). Brazilian archaeology from a brazilian perspective. *Antiquity*, 72:573–581.

Bartlett, M. S. (1974). The statistical analysis of spatial pattern. *Advances in Applied Probability*, 6(2):336–358.

Baxter, M. (2003). *Statistics in Archaeology*. Hodder Arnold, United Kingdom.

Beardsley, R. K. (1955). *Seminars in archaeology: 1955 : functional and evolutionary implications of community patterning*. Bobbs-Merrill.

Berger, A. L., Pietra, S. A. D., and Pietra, V. J. D. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics*, 22:39–71.

Berrocal, M. C. (2005). *Paisaje y arte rupestre: patrones de localización de la pintura levantina*, volume 1409 of *BAR International Series*. Archaeopress.

Besag, J. and Clifford, P. (1989). Generalized monte carlo significance tests. *Biometrika*, 76(4):633–642.

Besag, J. and Diggle, P. J. (1977). Simple monte carlo tests for spatial pattern. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26(3):327–333.

Bevan, A. and Conolly, J. (2006). Multiscalar approaches to settlement pattern analysis. In Lock, G. and Molyneaux, B. L., editors, *Confronting Scale in Archaeology*, pages 217–234. Springer.

Binford, L. R. (1972). *An archaeological perspective*. Studies in archeology. Seminar Press.

Binford, L. R. (1982). The archaeology of place. *Journal of Anthropological Archaeology*, 1(1):5 – 31.

Bivand, R., Pebesma, E., and Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*. Use R! Springer.

Blankholm, H. P. (2015). Macro-level predictive modeling of early post-glacial pioneer sites in varanger, norway. In Pizziolo, G. and Sarti, L., editors, *Predicting Prehistory - Predictive Models and Field Research Methods for Detecting Prehistoric Context*, number 11 in Millenni Studi di archeologia preistorica, pages 65–73. Museo e Instituto Fiorentino di Preistoria "Paolo Graziosi", Florence.

Blanton, R. E. (1976). Anthropological studies of cities. *Annual Review of Anthropology*, 5:249–264.

Bonham-Carter, G. F. (1994). *Geographic information systems for geoscientists: modelling with GIS*, volume 13 of *Computer Methods In the Geoscience*. Pergamon press.

Borsdorf, A. and Bender, O. (2010). *Allgemeine Siedlungsgeographie*. Böhlau Verlag, Wien, Köln, Weimar.

Botkin, D. B. (1990). *Discordant Harmonies: A New Ecology for the Twenty-First Century*. Oxford University Press, New York.

Bove, F. J. (1981). Trend surface analysis and the lowland classic maya collapse. *American Antiquity*, 46(1):93–112.

Bozarth, S. R., Price, K., Woods, W. I., Neves, E. G., and Rebellato, R. (2009). Phytoliths and terra preta: The hatahara site example. In Woods, W., Teixeira, W., Lehmann, J., Steiner, C., WinklerPrins, A., and Rebellato, L., editors, *Amazonian Dark Earths: Wim Sombroekś Vision*, pages 85–98. Springer Netherlands.

Brewster, A., Byrd, B. F., and Reddy, S. N. (2003). Cultural landscapes of coastal foragers: An example of gis and drainage catchment analysis from southern california. *Journal of GIS in Archaeology*, 1.

Browman, D. L. (1976). Demographic correlations of the wari conquest of junin. *American Antiquity*, 41(4):465–477.

Brown, K. L. (1979). Late prehistoric settlement patterns in southwest kansas. *Plains Anthropologist*, 24(85):191–206.

Butzer, K. W. (1982). *Archaeology as Human Ecology*. Cambridge University Press. Cambridge Books Online.

Carneiro, R. J. (1961). Slash-and-burn cultivation among the kuikuru and its implications for cultural development in the amazon basin. In Wilbert, J., editor, *The evolution of horticultural systems in native South America: Causes and Consequences*, number 2 in Antropológica Supplement Publication, pages 47–67. Sociedad de Ciencias Naturales La Salle, Caracas.

Carneiro, R. J. (1995). The history of ecological interpretations of amazonia: Does roosevelt have it right? In Sponsel, L. E., editor, *Indigenous Peoples and the Future of Amazonia*, pages 45–70. University of Arizona Press, Tucson.

Carneiro, R. L. (1970). A theory of the origin of the state: Traditional theories of state origins are considered and rejected in favor of a new ecological hypothesis. *Science*, 169(3947):733–738.

Carrión-Flores, C. and Irwin, E. G. (2004). Determinants of residential land-use conversion and sprawl at the rural-urban fringe. *American Journal of Agricultural Economics*, 86(4):889–904.

Carver, J. S. (1991). Integrating multi-criteria evaluation with geographical information systems. *International Journal of Geographical Information System*, 5(3):321–339.

Casarotto, A. (2015). Location preference modeling alias predictive modeling for the study of archaeological contexts. In Pizziolo, G. and Sarti, L., editors, *Predicting Prehistory - Predictive Models and Field Research Methods for Detecting Prehistoric Context*, number 11 in Millenni Studi di archeologia preistorica, pages 35–46. Museo e Instituto Fiorentino di Preistoria "Paolo Graziosi", Florence.

Cassels, R. (1972). Locational analysis of prehistoric settlement in new zealand. *Mankind*, 8(3):212–222.

Catana, A. J. (1963). The wandering quarter method of estimating population density. *Ecology*, 44(2):349–360.

Chang, K. C. (1972). *Settlement patterns in archaeology*. Addison-Wesley module in anthropology. Bibliography: p. 23-26.

Chawla, N. V. (2005). Data mining for imbalanced datasets: An overview. In Maimon, O. and Rokach, L., editors, *Data Mining and Knowledge Discovery Handbook*, pages 853–867. Springer.

Chisholm, M. (1973). *Rural settlement and land use*. Transaction Books.

Christaller, W. (1933). *Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Jena.

Christopherson, G. L., Barabe, P., and Johnson, P. S. (1999). Using arcgrid software's pathdistance function to model catchment zones for archaeological sites on the madaba plain, jordan. In *Proceedings esri User Conferenece 1999*, San Diego, California.

Clapham, A. R. (1936). Over-dispersion in grassland communities and the use of statistical methods in plant ecology. *Journal of Ecology*, 24(1):pp. 232–251.

Clark, G. (1957). *Archaeology And Society*. Methuen & Co, 3rd edition.

Clark, G. A., Locht, J.-L., Séara, F., Rotillon, S., Cupillard, C., Eriksen, B. V., Bratlund, B., de Bie, M., and Caspar, J.-P. (2005). Modern approaches to paleolithic archaeology in Europe – a sampler of research traditions. *American Antiquity*, 70(2):376–384.

Clark, P. J. and Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35(4):445–453.

Clarke, D. L. (1977). *Spatial archaeology*. Academic Press.

Clarke, R. D. (1946). An application of the poisson distribution. *Journal of the Institute of Actuaries*, 72:481.

Cliff, A. D. and Ord, J. K. (1973). *Spatial autocorrelation*. Monographs in spatial and environmental systems analysis. Pion.

Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes : Models and Applications*. Pion, London.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

Connell, J. H. (1978). Diversity in tropical rain forests and coral reefs. *Science*, 199(4335):1302–1310.

Conolly, J. and Lake, M. (2006). *Geographical information systems in archaeology*. Cambridge University Press.

Cormack, R. M. (1979). Spatial aspects of competition between individuals. In Cormack, R. M. and Ord, J. K., editors, *Spatial and Temporal Analysis in Ecology*, pages 115–211. Fairland: International Co-operatice Publishing House.

Corrêa, C. G. (1987). Horticultores pré-históricos do litoral do pará, brasil. *Revista de Arqueologia*, 4(20):137–252.

Cottam, G. and Curtis, J. T. (1949). A method for making rapid surveys of woodlands by means of pairs of randomly selected trees. *Ecology*, 30(1):101–104.

Countryman, J., Carrier, S., and Kane, S. (2010). Predictive modelling: a case study of agricultural terraces at monte pallano (abruzzo, italy). In Zhou, M., Romanowska, I., Wu, Z., Xu, P., and Verhagen, P., editors, *Revive the Past. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 39th International Conference, Beijing, April 12-16*, pages 307–317. Pallas Publications.

Cox, T. F. (1981). Reflexive nearest neighbours. *Biometrics*, 37(2):pp. 367–369.

Cressie, N. (1993). *Statistics for spatial data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. John Wiley & Sons.

da Costa, A. C. S. and Nanni, M. R. (2006). Classification systems: Brazilian. encyclopedia of soil science. In Lal, R., editor, *Encyclopedia of Soil Science*, volume 1, pages 234–238. CRC Press, New York.

Dacey, M. F. (1963). Order neighbor statstics for a class of random patterns in multidimensional space. *Annals of the Association of American Geographers*, 53(4):505–515.

Dalton, G. (1969). Theoretical issues in economic anthropology. *Current Anthropology*, pages 63–102.

Danks, P. (1977). Some observations on medieval and post- medieval artifact distributions: A spatial model at the regional scale (macro). In Clarke, D. L., editor, *Spatial Archaology*. Academic Press.

Darroch, J. N. and Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, 43(5):1470–1480.

Davey, P. J. (1971). The distribution of later bronze age metalwork from lincolnshire. *Proceedings of the Prehistoric Society*, 37(01):96–111.

David, F. N. and Moore, P. G. (1954). Notes on contagious distributions in plant populations. *Annals of Botany*, 18(1):47–53.

Davis, J. C. (1973). *Statistics and Data Analysis in Geology*. John Wiley & Sons, Inc., New York.

De Carvajal, F. G. (1942). *Relación del nuevo descubrimiento del famoso Río Grande que descubrió por muy gran ventura el capitán Francisco de Orellana*. Ministerio de Educación.

De Silva, M. and Pizziolo, G. (2001). Setting up a "human calibrated" anisotropic cost surface for archaeological landscape investigation. In Stančič, Z. and Veljanovski, T., editors, *Computing Archaeology for Understanding the Past. CAA 2000. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 28th Conference, Ljubljana, April 2000*, number 931 in BAR International Series, pages 279–286, Oxford. Archaeopress.

de Smith, M. J. (2006). Determination of gradient and curvature constrained optimal paths. *Computer-Aided Civil and Infrastructure Engineering*, 21(1):24–38.

Delamater, P. L., Messina, J. P., Shortridge, A. M., and Grady, S. C. (2012). Measuring geographic access to health care: raster and network-based methods. *International Journal of Health Geographics*, 11(1):1–18.

Della Pietra, S., Della Pietra, V., and Lafferty, J. (1997). Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4):380–393.

Denevan, W. M. (1966). A cultural-ecological view of the former aboriginal settlement in the amazon basin. *The Professional Geographer*, 18(6):346–351.

Denevan, W. M. (1976). The aboriginal population of western amazonia in relation to habitat and subsistence. In Denevan, W. M., editor, *The Native Population of the Americas in 1492*, pages 205–234. University of Wisconsin Press, Madison.

Denevan, W. M. (1996). A bluff model of riverine settlement in prehistoric amazonia. *Annals of the Association of American Geographers*, 86(4):654–681.

Denevan, W. M. (2001). *Cultivated landscapes of native Amazonia and the Andes*. Oxford University Press.

Deravignone, L., Blankholm, H. P., and Pizziolo, G. (2015). Predictive modeling and artificial neural networks: From model to survey. In *Mathematics and Archaeology*, pages 335–351. CRC Press.

Deravignone, L. and Macchi Jánica, G. (2006a). Artificial neural networks in archaeology. *Archeologia e Calcolatori*, (XVII):121–136.

Deravignone, L. and Macchi Jánica, G. (2006b). From gis to artificial neural networks: ten year studies on fortified villages in central italy. In *Archäologie und Computer 2005, Workshop 10*, Vienna.

Diamond, J. and Bellwood, P. (2003). Farmers and their languages: The first expansions. *Science*, 300(5619):597–603.

Dibyopama, A. (2010). Site catchment analysis of balathal. *Ancient Asia*, 2:47–57.

Diggle, P. J. (2003). *Statistical analysis of spatial point patterns.* Edward Arnold. 2nd edition.

Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. Chapman and Hall/CRC, 3rd edition.

Dixon, P. M. (2002). Ripley's k function. *Encyclopedia of environmetrics*, 3:1796–1803.

Dixon, P. M. (2012). *Nearest Neighbor Methods*. John Wiley & Sons, Ltd.

Douglas, J. B. (1975). Clustering and aggregation. *Sankhyã: The Indian Journal of Statistics, Series B (1960-2002)*, 37(4):398–417.

Drennan, R. D. and Peterson, C. E. (2004). Comparing archaeological settlement systems with rank-size graphs: a measure of shape and statistical confidence. *Journal of Archaeological Science*, 31(5):533–549.

Drury, P. J. (1972). Preliminary report on the romano-british settlement at chelmsford, essex. *Essex Archaeology and History*, 4:3–29.

Du Rietz, G. (1929). *The Fundamental Units of Vegetation*.

Ducke, B. and Kroefges, P. C. (2007). From points to areas: constructing territories from archaeological site patterns using an enhanced xtent model. In *Computer Applications and Quantitative Methods in Archaeology 2007*.

Duffy, P. R. (2010). *Complexity and Autonomy in Bronze Age Europe: Assessing Cultural Developments in the Eastern Hungary*. PhD thesis, University of Michigan.

Egeland, C., Nicholson, C., and Gasparian, B. (2010). Using gis and ecological variables to identify high potential areas for paleoanthropological survey: An example from northern armenia. *Journal of Ecological Anthropology*, 14(1):89–98.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., and E. Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2):129–51.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., and Yates, C. J. (2011). A statistical explanation of maxent for ecologists. *Diversity and Distributions*, 17(1):43–57.

Ellison, A. and Harriss, J. (1972). Settlement and land use in the prehistory and early history of southern england: a study based on locational models. In Clarke, D. L., editor, *Models in archaeology*, pages 911–962. Methuen, London.

Erickson, C. L. (2000). An artificial landscape-scale fishery in the bolivian amazon. *Nature*, 408(6809):190–193.

Eriksen, L. (2011). *Nature and Culture in Prehistoric Amazonia: Using GIS to Reconstruct Ancient Ethnogenetic Processes from Archeology, Linguistics, Geography, and Ethnohistory*. PhD thesis, Department of Human Geography, Human Ecology Division, Lund University.

Estivill-Castro, V. (2002). Why so many clustering algorithms: A position paper. *SIGKDD Explorations Newsletter*, 4(1):65–75.

Estivill-Castro, V. and Yang, J. (2000). Fast and robust general purpose clustering algorithms. In Mizoguchi, R. and Slaney, J., editors, *PRICAI 2000 Topics in Artificial Intelligence*, volume 1886 of *Lecture Notes in Computer Science*, pages 208–218. Springer, Berlin Heidelberg.

Evans, C. and Meggers, B. J. (1973). United states "imperialism" and latin american archaeology. *American Antiquity*, 38(3):257–258.

Eve, S. J. and Crema, E. R. (2014). A house with a view? multi-model inference, visibility fields, and point process analysis of a bronze age settlement on leskernick hill (cornwall, uk). *Journal of Archaeological Science*, 43:267–277.

Fagerlund, V. G. and Smith, R. H. T. (1970). A preliminary map of market periodicities in ghana. *The Journal of Developing Areas*, 4(3):333–348.

Fan, F., Wang, Y., Qiu, M., and Wang, Z. (2009). Evaluating the temporal and spatial urban expansion patterns of guangzhou from 1979 to 2003 by remote sensing and GIS methods. *International Journal of Geographical Information Science*, 23(11):1371–1388.

Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P. (1996a). From data mining to knowledge discovery – an overview. In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., editors, *Advances in knowledge discovery and data mining*, pages 1–30. American Association for Artificial Intelligence, Menlo Park.

Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P. (1996b). From data mining to knowledge discovery in databases. *American Association for Artificial Intelligence*, 17(3).

Field, J. S., Petraglia, M. D., and Lahr, M. M. (2007). The southern dispersal hypothesis and the south Asian archaeological record: Examination of dispersal routes through gis analysis. *Journal of Anthropological Archaeology*, 26(1):88–108.

Findlow, F. J. and DeAtley, S. P. (1974). Prehistoric land use patterns in the animas valley: A first approximation. In *Anthropology UCLA*, volume 6, pages 1–57.

Fisher, R. A., Thornton, H. G., and MacKenzie, W. A. (1922). The accuracy of the plating method of estimating the densityn of bacterial populations. *Annals of Applied Biology*, 9(3-4):325–359.

Flannery, K. V. (1976). Empirical determination of site catchments in oaxaca and tehuacan. In Flannery, K. V., editor, *The early Mesoamerican village*, pages 103–117. Academic Press, New York.

Fortes, M., Steel, R. W., and Ady, P. (1947). Ashanti survey, 1945-46: An experiment in social research. *The Geographical Journal*, 110(4/6):pp. 149–177.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Quantitative Geography*. SAGE Publications, London.

Fowlkes, E. B. and Mallows, C. L. (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, 78(383):553–569.

Franklin, J. (2009). *Mapping species distributions: Spatial Inference and Prediction*. Cambridge University Press.

Funari, P. P. A. (1995). Mixed features of archaeological theory in brazil. In Ucko, P. J., editor, *Theory in Archaeology: A World Perspective*, pages 232–246. Routledge.

Garcia, J. M. V. (2015). Theoretical remarks on predictive models in landscape archaeology. In Pizziolo, G. and Sarti, L., editors, *Predicting Prehistory - Predictive Models and Field Research Methods for Detecting Prehistoric Context*, number 11 in Millenni Studi di archeologia preistorica, pages 25–33. Museo e Instituto Fiorentino di Preistoria "Paolo Graziosi", Florence.

Garner, B. J. (1967). Models of urban geography and settlement location. In Chorley, R. J. and Haggett, P., editors, *Models in geography*. Methuen & Co.

Getis, A. (1964). Temporal land-use pattern analysis with the use of nearest neighbor and quadrat methods. *Annals of the Association of American Geographers*, 54(3):391–399.

Getis, A. and Franklin, J. (1987). Second-order neighborhood analysis of mapped point patterns. *Ecology*, 68(3):473–477.

Gilman, A. and Thornes, J. B. (1985). *Land-use and Prehistory in South-East Spain*. George Allen and Unwin, London.

Glaser, B., Zech, W., and Woods, W. (2004). History, current knowledge and future perspectives of geoecological research concerning the origin of amazonian anthropogenic dark earths (terra preta). In Glaser, B. and Woods, W. I., editors, *Amazonian Dark Earths: Explorations in Space and Time*, pages 9–17. Springer, Berlin Heidelberg.

Glass, C., Steele, J., and Wheatley, D. (1999). Modelling spatial range expansion across a heterogeneous cost surface. In *Proceedings CAA 97, Birmingham*, pages 67–72.

Goebel, A. (2007). Sustainable urban development? low-cost housing challenges in south africa. *Habitat International*, 31(3-4):291–302.

Goggin, J. M. (1949). Cultural traditions in florida prehistory. In Griffin, J. W., editor, *The Florida Indian and His Neighbors*, pages 13–44. Rollins College Inter-American Center, Winter Park, Florida.

Gonzalez-Abraham, C., Radeloff, V., Hammer, R., Hawbaker, T., Stewart, S., and Clayton, M. (2007). Building patterns and landscape fragmentation in northern wisconsin, usa. *Landscape Ecology*, 22(2):217–230.

Goodall, D. W. (1952). Quantitative aspects of plant distribution. *Biological Reviews*, 27(2):194–242.

Goodall, D. W. (1970). Statistical plant ecology. *Annual Review of Ecology and Systematics*, 1(1):99–124.

Grady, J. (1980). *Environmental FFactor in Archaeological Site Locations. PiceanceBasins, Colorado*. Number 9 in Cultural Resources Series. Bureau of Land Management, Colorado.

Graham, C. and Talay, D. (2013). *Stochastic Simulation and Monte Carlo Methods*. Springer Berlin Heidelberg.

Grant, E. (1986). Hill-forts, central places and territories. In Grant, E., editor, *Central places, archaeology and history*, pages 13–32. Department of Archaeology and Prehistory, University of Sheffield.

Green, E. L. (1973). Location analysis of prehistoric maya sites in northern british honduras. *American Antiquity*, 38(3):279–293.

Greig-Smith, P. (1952). The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany*, 16:293–316.

Greig-Smith, P. (1983). *Quantitative plant ecology*, volume 9 of *Studies in Ecology*. University of California Press, 3rd edition.

Guo, D. and Mennis, J. (2009). Spatial data mining and geographic knowledge discovery – an introduction. *Computers, Environment and Urban Systems*, 33(6):403–408.

Guo, D., Peuquet, D., and Gahegan, M. (2003). Iceage: Interactive clustering and exploration of large and high-dimensional geodata. *GeoInformatica*, 7(3):229–253.

Haggett, P. (2004). *Geographie - Eine globale Synthese*. UTB, Stuttgart, 3rd edition.

Hambrick, D. C. (1983). An empirical typology of mature industrial-product environments. *Academy of Management Journal*, 26(2):213–230.

Hames, R. B. and Vickers, W. T. (1983). Chapter 1 - introduction. In *Adaptive Responses of Native Amazonians*, pages 1–26. Academic Press, San Diego.

Han, J., Kamber, M., and Tung, A. K. H. (2001). Spatial clustering methods in data mining: a survey. *Data Mining and Knowledge Discovery*.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36.

Hare, T. S. (2001). *Political Economy, Spatial Analysis and Postclassic Sites in the Yautepec Valley, Mexico*. PhD thesis, The University at Albany, State University of New York.

Hartt, C. F. (1885). *Contribuições para a ethnologia do Valle do Amazonas*, volume 6. Archivos do Museu Nacional do Rio de Janeiro.

Harvey, D. (1968). Some methodological problems in the use of the neyman type a and the negative binomial probability distributions for the analysis of spatial point patterns. *Transactions of the Institute of British Geographers*, (44):85–95.

Harvey, D. (1969). *Explanation in Geography*. Edward Arnold.

Heckbert, S. (2013). Mayasim: An agent-based model of the ancient maya social-ecological system. *Journal of Artificial Societies and Social Simulation*, 16(4):11.

Heckenberger, M. (2008). Amazonian mosaics: Identity, interaction, and integration in the tropical forest. In Silverman and Isbell (2008), pages 941–962.

Heckenberger, M. and Neves, E. G. (2009). Amazonian archaeology. *Annual Review of Anthropology*, 38(1):251–266.

Heckenberger, M. J. (1998). De onde surgem os modelos? as origens e expansões tupi na amazânia central. *Revista de Antropologia*, 41(1):69–96.

Heckenberger, M. J., Petersen, J. B., and Neves, E. G. (1999). Village size and permanence in amazonia: Two archaeological examples from brazil. *Latin American Antiquity*, 10(4):353–376.

Heckenberger, M. J., Russell, J. C., Fausto, C., Toney, J. R., Schmidt, M. J., Pereira, E., Franchetto, B., and Kuikuro, A. (2008). Pre-columbian urbanism, anthropogenic landscapes, and the future of the amazon. *Science*, 321(5893):1214–1217.

Hempel, C. G. (1970). *Aspects of scientific explanation, and other essays in the philosophy of science*. Free Press, New York.

Henshall, J. D. (1968). Models of agricultural activity. *Socio-economic models in geography*, pages 425–458.

Higgs, E. S. (1975). *Palaeoeconomy*. Cambridge University Press.

Higgs, E. S. and Vita-Finzi, C. (1972). Prehistoric economies: A territorial approach. In Higgs, E. S., editor, *Papers in Economic Prehisty*. Cambridge University Press.

Hilbert, P. P. (1968). *Archäologische Untersuchungen am mittleren Amazonas. Beiträge zur Vorgeschichte des südamerikanischen Tieflandes*. Marburger Studien zur Völkerkunde. Bd. 1.

Hilbert, P. P. (1977). Forschungen in der amazonas-region zwischen anden und atlantik. In Frey, O.-H. and Roth, H., editors, *Kleine Schriften aus dem Vorgeschichtlichen Seminar Marburg vom 27. VI. 1977 bis zum 3. VII. 1977*, volume 2, page 12. Marburg Universität.

Hilbert, P. P. and Hilbert, K. (1980). Resultados preliminares da pesquisa arqueológica nos rios nhamundáe trombetas, baixo rio amazonas. *Antropologica*, 2(75):1–15.

Hill, P. and Smith, R. H. T. (1972). The spatial and temporal synchronization of periodic markets: Evidence from four emirates in northern nigeria. *Economic Geography*, 48(3):345–355.

Hodder, I. (1971). The use of nearest-neighbour analysis. *Cornish Archaeology*, 10:35–36.

Hodder, I. (1977). Some new directions in the spatial analysis of archaeological data at the regional scale. In Clarke, D. L., editor, *Spatial Archaeology*, chapter 6. Academic Press.

Hodder, I. and Hassall, M. (1971). The non-random spacing of romano-british walled towns. *Man*, 6(3):391–407.

Hodder, I. and Orton, C. (1976). *Spatial analysis in archaeology*. Number 1 in New studies in archaeology. Cambridge University Press, Cambridge.

Holgate, P. (1965). Some new tests of randomness. *Journal of Ecology*, 53(2):261–266.

Hope, A. C. A. (1968). A simplified monte carlo significance test procedure. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 582–598.

Hornborg, A. (2005). Ethnogenesis, regional integration, and ecology in prehistoric amazonia. *Current Anthropology*, 46(4):589–620.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218.

Hudson, J. C. (1969). A location theory for rural settlement. *Annals of the Association of American Geographers*, 59(2):365–381.

Huerta Munoz, U. and Källestâl, C. (2012). Geographical accessibility and spatial coverage modeling of the primary health care network in the western province of Rwanda. *Int J Health Geogr*, 11(1):40.

Hurlbett, R. E. (1977). *Environmental Constraint and Settlement Predictability, Northwestern Colorado*. Number 3 in Cultural Resource Series. Department of the Interior, Bureau of Land Management, Colorado State Office, Denver, Colorado.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3):264–323.

Janssen, R. and Rietveld, P. (1990). Multicriteria analysis and geographical information systems: an application to agricultural land use in the netherlands. *Geographical information systems for urban and regional planning*, 17:129–139.

Jarman, M. R. (1972). A territorial model for archaeology: A behavioural and geoarchaeological approach. In Clarke, D. L., editor, *Models in Archaeology*, pages 705–734. Methuen, London.

Jaynes, E. (1991). Notes on present status and future prospects. In Grandy, W. T., J. and Schick, L. H., editors, *Maximum Entropy and Bayesian Methods*, volume 43 of *Fundamental Theories of Physics*, pages 1–13. Springer Netherlands.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106:620–630.

Jiménez-Valverde, A., Lobo, J. M., and Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, 14(6):885–890.

Jochim, M. (1976). *Hunter-gatherer Susbsistence and Settlement: A Predictive Model*. Studies in archeology series. Academic Press.

Johnson, G. A. (1977). Aspects of regional analysis in archaeology. *Annual Review of Anthropology*, 6:479–508.

Judge, W. J. (1973). *Paleoindian Occupation of the Central Rio Grande Valley in New Mexico*. University of New Mexico Press, Albuquerque.

Kabacoff, R. I. (2011). *R in Action*. Manning Publications Co.

Kamermans, H., van Leusen, P. M., and Verhagen, P. (2009). *Archaeological Prediction and Risk Management. Alternatives to current practice*. Number 17 in Archaeological Studies Leiden University Press Series. Leiden University Press.

Kamermans, H. and Wansleeben, M. (1999). Predictive modelling in dutch archaeology, joining forces. In Barceló, J. A., Briz, I., and Vila, A., editors, *New Techniques for Old Times. CAA98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998*, volume 757 of *BAR International Series*, pages 225–230, Oxford. Archaeopress.

Kämpf, N., Woods, W. I., Sombroek, W., Kern, D. C., and Cunha, T. J. (2003). Classification of amazonian dark earths and other ancient anthropic soils. In Lehmann, J., Kern, D., Glaser, B., and Wodos, W., editors, *Amazonian Dark Earths*, pages 77–102. Springer Netherlands.

Kantner, J. (2008). The archaeology of regions: From discrete analytical toolkit to ubiquitous spatial perspective. *Journal of Archaeological Research*, 16(1):37–81.

Kariel, H. G. (1970). Analysis of the alberta settlement pattern for 1961 and 1966 by nearest neighbor analysis. *Geografiska Annaler. Series B, Human Geography*, 52(2):124–130.

Katzer, F. (1903). *Grundzüge der Geologie des unteren Amazonasgebietes (des Staates ParÃ¡ in Brasilien)*. Max Weg.

Kaufman, L. and Rousseeuw, P. J. (1990). Finding groups in data. *Wiley Series in Probability and Statistics*.

Kern, D. C., da Costa, M. L., and Frazão, F. J. L. (2004). Evolution of the scientific knowledge regarding archaeological black earths of amazonia. In Glaser, B. and Woods, W. I., editors, *Amazonian Dark Earths: Explorations in Space and Time*, pages 19–28. Springer Berlin Heidelberg.

Kern, D. C., D'aquino, G., Rodrigues, T. E., Frazao, F. J. L., Sombroek, W., Myers, T. P., and Neves, E. G. (2003). Distribution of amazonian dark earths in the brazilian amazon. In Lehmann, J., Kern, D. C., Glaser, B., and Wodos, W. I., editors, *Amazonian Dark Earths*, pages 51–75. Springer Netherlands.

Ketchen, D. J. and Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, 17(6):441–458.

Ketchen, D. J., Thomas, J. B., and Snow, C. C. (1993). Organizational configurations and performance: A comparison of theoretical approaches. *The Academy of Management Journal*, 36(6):1278–1313.

Kidder, A. V., Shook, E. M., and Jennings, J. D. (1946). *Excavations at Kaminaljuyu, Guatemala*. Carnegie Institution of Washington Washington, D.C.

Kincaid, C. (1988). Predictive modeling and its relationship to cultural resource management applications. In Judge, J. W. and Sebastian, L., editors, *Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modelling*, pages 549–569. U.S. Department of the Interior, Bureau of Land Management, Denver.

King, L. J. (1968). A quantitative expression of the pattern of urban settlements in selected areas of the united states. In Berry, B. J. L. and Marble, D. F., editors, *Spatial Analysis - A Reader in Statistical Geography*. Prenti.

King, L. J. (1969). *Statistical analysis in geography*. Prentice-Hall.

Kirch, P. V., Hartshorn, A. S., Chadwick, O. A., Vitousek, P. M., Sherrod, D. R., Coil, J., Holm, L., and Sharp, W. D. (2004). Environment, agriculture, and settlement patterns in a marginal polynesian landscape. *Proceedings of the National Academy of Sciences*, 101(26):9936–9941.

Koch-Grünberg, T. (1923). *Vom Roroima zum Orinoco - Ergebnisse einer Reise in Nordbrasilien und Venezuela in den Jahren 1911-1913 Band 3*. Verlag Strecker und Schröder.

Kohler, T. A. (1988). Predictive modelling: History and practice. In Judge, J. W. and Sebastian, L., editors, *Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modelling*, pages 19–59. U.S. Department of the Interior, Bureau of Land Management, Denver.

Kohler, T. A. and Parker, S. C. (1986). Predictive models for archaeological resource location. *Advances in Archaeological Method and Theory*, 9:397–452.

Kowalewski, S. A. (2008). Regional settlement pattern studies. *Journal of Archaeological Research*, 16(3):225–285.

Kreyszig, E. (1968). *Statistische Methoden und ihre Anwendungen*. Vandenhoeck & Ruprecht, Göttingen, 3rd edition.

Kroeber, A. (1944). *Peruvian Archeology in 1942*. Number 4 in Viking Fund publications in anthropology. Wenner Gren.

Kruk, J. and Machnik, J. (1973). *Studia osadnicze nad neolitem wyżyn lessowych*. Ossolineum.

Kvamme, K. L. (1980). Predictive model of site location in the glenwood springs resource area. In et. al., B. R. J., editor, *Class II cultural resource inventory of the Glenwood Springs resource area, Grand Junction District, Colorado*, pages 88–136. Bureau of Land Management, Grand Junction District, Colorado.

Kvamme, K. L. (1983). *New Methods for investigating the environmental basis of prehistoric site locations*. PhD thesis, University of California, Santa Barbara.

Kvamme, K. L. (1985). Determining empirical relationships between the natural environment and prehistoric site locations: a hunter-gatherer example. In Carr, C., editor, *For concordance in archaeological analysis (C. Carr, ed.). Westport Publishers, Inc., Kansas City, Missouri*, pages 208–238. Westport Press, Kansas City.

Kvamme, K. L. (1988). Development and testing of quantitative models. In Judge, J. W. and Sebastian, L., editors, *Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modelling*, chapter 8, pages 325–428. U.S. Department of the Interior, Bureau of Land Management, Denver.

Kvamme, K. L. (1997). Ranters corner bringing the camps together: Gis and ed. *Archaeological Computing Newsletter*, 1(47):1–4.

Kvamme, K. L. and Kohler, T. A. (1988). Geographic information systems: Technical aids for data collection, analsyis, and display. In Judge, J. W. and Sebastian, L., editors, *Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modelling*, chapter 10, pages 493–547. U.S. Department of the Interior, Bureau of Land Management, Denver.

Lathrap, D. W. (1970). *The Upper Amazon. Ancient peoples and places*. Thames & Hudson Ltd.

Lathrap, D. W. (1973a). The antiquity and importance of long-distance trade relationships in the moist tropics of pre-columbian south america. *World Archaeology*, 5(2):170–186.

Lathrap, D. W. (1973b). Gifts of the cayman: Some thoughts on the subsistence basis of chavin. In Lathrap, D. W. and Douglas, J., editors, *Variation In Anthropoology: Essays In Honor of John C. McGregor*, pages 91–105. Illinois Archaeological Survey.

Lathrap, D. W. (1974). The moist tropics, the arid lands, and the appearance of great art styles in the new world. art and the environment in native america. In King, M. E. and Traylor, J. I. R., editors, *Special Publications*, volume 7, pages 115–158. Museum of Texas Tech University, Lubbock.

Lee, R. B. (1969). !kung bushman subsistence: An input-output analysis. In Vayda, A. P., editor, *Environment and cultural behavior*. Natural History Press, Garden City, New York.

Leonardi, G. (1992). Assunzione e analisi dei dati territoriali in funzione della valutazione della diacronia e delle modalità del popolamento. In Bernardi, M., editor, *Archeologia del paesaggio*, volume 4 of *Ciclo di lezioni sulla ricerca applicata in archeologia*.

Lewis, R. (1996). *Kentucky Archaeology*. Perspectives on Kentucky's past. University Press of Kentucky.

Li, Y., Liu, C., Zhang, H., and Gao, X. (2011). Evaluation on the human settlements environment suitability in the three gorges reservoir area of chongqing based on rs and gis. *Journal of Geographical Sciences*, 21(2):346–358.

Linard, C., Gilbert, M., Snow, R. W., Noor, A. M., and Tatem, A. J. (2012). Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS ONE*, 7(2):1–8.

Liu, J., Zhan, J., and Deng, X. (2005). Spatio-temporal patterns and driving forces of urban land expansion in china during the economic reform era. *Ambio*, pages 450–455.

Lloyd, M. (1967). Mean crowding. *Journal of Animal Ecology*, 36(1):1–30.

Lönnqvist, M., Törmä, M., Lönnqvist, K., and Nueõz, M. (2009). Site catchment analysis applied to the pastoral landscape of jebel bishri in syria by using quickbird satellite imagery and aster-dem data. In *XXII CIPA Symposium Kyoto, Japan 11 October - 15 October 2009 Proceedings*.

Lovis, J. W. A. (1976). Quarter sections and forests: An example of probability sampling in the northeastern woodlands. *American Antiquity*, 41(3):364–372.

Lowie, R. H. (1948). The tropical forest tribes. an introduction. In Steward (1947), pages 1–56. 1940-1947.

Luck, G. W. (2007). The relationships between net primary productivity, human population density and species conservation. *Journal of Biogeography*, 34(2):201–212.

Machado, J. S. (2005). *Montículos Artificiais na Amazônia Central: Um estudo de caso do sítio Hatahara*. PhD thesis, Universidade de São Paulo.

Maimon, O. and Rokach, L. (2006a). *Data Mining and Knowledge Discovery Handbook*. Springer.

Maimon, O. and Rokach, L. (2006b). Introduction to knowledge discovery in databases. In Maimon and Rokach (2006a), chapter 1.

Manly, B. F. J. (2004). *Multivariate Statistical Methods: A Primer, Third Edition*. CRC Press, 3rd edition.

Mann, C. C. (2002). 1491. *Atlantic Monthly*, 289(3):41–53.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate analysis*. Academic press.

Mayer, C. (2006). Making use of distances: Estimating parameters of spatial processes. In Mehrer, M. W. and Wescott, K. L., editors, *GIS and Archaeological Site Location Modeling*, pages 137–153. CRC Press.

McGregor, J. C. (1950). Weighted traits and traditions. In Reed, E. K. and King, D. S., editors, *For the Dean*, pages 291–198. Hohokam Museum Association and Southwestern Monuments Association.

McKelvey, B. (1975). Guidelines for the empirical classification of organizations. *Administrative Science Quarterly*, 20(4):509–525.

McKelvey, B. (1978). Organizational systematics: Taxonomic lessons from biology. *Management Science*, 24(13):1428–1440.

McMichael, C. H., Palace, M. W., Bush, M. B., Braswell, B., Hagen, S., Neves, E. G., Silman, M. R., Tamanaha, E. K., and Czarnecki, C. (2014). Predicting pre-columbian anthropogenic soils in amazonia. *Proceedings of the Royal Society B: Biological Sciences*, 281(1777).

Meggers, B. (1984). The indigenous peoples of amazonia, their cultures, land use patterns and effects on the landscape and biota. In Sioli, H., editor, *The Amazon*, volume 56 of *Monographiae Biologicae*, pages 627–648. Springer Netherlands.

Meggers, B. J. (1971). *Amazonia: Man and Culture in a Counterfeit Paradise*. Harlan Davidson.

Meggers, B. J. (1979). Climatic oscillation as a factor in the prehistory of amazonia. *American Antiquity*, 44(2):252–266.

Meggers, B. J. (1987a). The early history of man in amazonia. In Whitmore, T. C. and Prance, G. T., editors, *Biogeography and Quaternary History in Tropical America*, pages 151–174. Clarendon Press, Oxford.

Meggers, B. J. (1987b). El uso de secuencias ceramicas seriadas para inferir conducta social. In *Actas del Primer Simposio de la Fundacion de Arqueologic del Caribe*, pages 11–32. Smithsonian Institution Press.

Meggers, B. J. (1988). Identification and implications of a hiatus in the archaeological sequence on marajó island, brazil. *Journal of the Washington Academy of Sciences*, 78(3):145–153.

Meggers, B. J. and Evans, C. (1957). *Archeological investigations at the mouth of the Amazon*. Smithsonian Institution.

Meggers, B. J. and Evans, C. (1961). An experimental formulation of horizon styles in the tropical forest area of south america. In Lothrop, S. K., editor, *Essays in pre-columbian art and archaeology*, pages 46–82. Harvard University Press, Cambridge.

Meggers, B. J. and Evans, C. (1983). Lowland south america and the antilles. In Jennings, J. D., editor, *Ancient Native Americans*, pages 287–335. W. H. Freeman and Co, San Francisco.

Meilă, M. (2007). Comparing clusterings – an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895.

Meilă, M. and Heckerman, D. (2001). An experimental comparison of model-based clustering methods. *Machine Learning*, 42(1-2):9–29.

Mennis, J. and Liu, J. W. (2005). Mining association rules in spatio-temporal data: An analysis of urban socioeconomic and land cover change. *Transactions in GIS*, 9(1):5–17.

Menze, B. H. and Ur, J. A. (2012). Mapping patterns of long-term settlement in northern mesopotamia at a large scale. *Proceedings of the National Academy of Sciences*, 109(14):778–787.

Merow, C., Smith, M. J., and Silander, J. A. (2013). A practical guide to maxent for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36(10):1058–1069.

Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–341.

Miller, E. T. (1992). *Arqueologia nos empreendimentos hidrelétricos da Eletronorte: resultados preliminares*. Centrais Elétricas do Norte do Brasil, Brasília.

Miller, E. T. and Caldarelli, S. B. (1987). *Estudos de viabilidade arqueológica nas áreas de inundação e de influência da UHE Ji-Paraná. Relatório de atividades.* Museu Paraense Emílio Goeldi, Belém.

Miller, H. J. and Han, J. (2003). Geographic data mining and knowledge discovery – an overview. In Miller, H. J. and Han, J., editors, *Geographic Data Mining and Knowledge Discovery*. CRC Press.

Milligan, G. W. and Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179.

Modenesi, M. V., Costa, M. C. A., Evsukoff, A. G., and Ebecken, N. F. F. (2007). Parallel fuzzy c-means cluster analysis. In Daydé, M., Palma, J. M. L. M., Coutinho, Á. L. G. A., Pacitti, E., and Lopes, J. C., editors, *High Performance Computing for Computational Science - VECPAR 2006*, volume 4395 of *Lecture Notes in Computer Science*, pages 52–65. Springer Berlin Heidelberg.

Moellering, H. and Tobler, W. (1972). Geographical variances. *Geographical Analysis*, 4(1):34–50.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: an evaluation. *The Computer Journal*, 20(4):359–363.

Monroe, C. B., Feder, K. L., Poirier, D. A., and Gradie, R. R. (1980). Trend surface analysis: the settlement dynamics of colonial connecticut. *Man in the Northeast*, 19:105–128.

Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, 52(1-2):91–118.

Moran, E. F. (1993). *Through Amazonian Eyes: The Human Ecology of Amazonian Populations*. University of Iowa Press.

Moran, E. F. (1995). Disaggregating amazonia: A strategy for understanding biological and cultural diversity. In Sponsel, L. E., editor, *Indigenous Peoples and the Future of Amazonia*, pages 71–95. University of Arizona Press, Tucson.

Morgan, L. H. (1881). *Houses and House-life of the American Aborigines*. United States Government Publishing Office.

Morisita, M. (1959). Measuring of the dispersion and analysis of distribution patterns. *Memoires of the Faculty of Science, Kyushu University, Series E. Biology*, 2:215–235.

Mueser, P. R. and Graves, P. E. (1995). Examining the role of economic opportunity and amenities in explaining population redistribution. *Journal of Urban Economics*, 37(2):176–200.

Murphy, A. H. and Winkler, R. L. (1992). Diagnostic verification of probability forecasts. *International Journal of Forecasting*, 7(4):435–455.

Myers, T. P. (2004). Dark earth in the upper amazon. In Glaser, B. and Woods, W. I., editors, *Amazonian Dark Earths: Explorations in Space and Time*, pages 67–94. Springer Berlin Heidelberg.

Neves, E. G. (1998a). *Paths in Dark Waters: Archaeology as Indigenous History in the Upper Rio Negro Basin, Northwest Amazon*. PhD thesis, Indiana University.

Neves, E. G. (1998b). Twenty years of amazonian archaeology in brazil (1977 - 1997). *Antiquity*, 72:625–632.

Neves, E. G. (1999). Changin perspective in amazonian archaeology. In Alberti, B. and Politis, G., editors, *Archaeology in Latin America*, pages 219–249. Routledge.

Neves, E. G. (2008). *Ecology, Ceramic Chronology and Distribution, Long-term History, and Political Change in the Amazonian Floodplain*, chapter 20, pages 359–379. In Silverman and Isbell (2008).

Neves, E. G., Petersen, J. B., Bartone, R. N., and Augusto Da Silva, C. (2003). Historical and socio-cultural origins of amazonian dark earth. In Lehmann, J., Kern, D. C., Glaser, B., and Woods, W. I., editors, *Amazonian Dark Earths*, pages 29–50. Springer Netherlands.

Nimuendaju, C. (1952). The tapajós. *Kroeber Anthropological Society Papers*, 6:1–25.

Norton, Melanie, J. (1999). Knowledge discovery in databases. In Qin, J. and Norton, M. J., editors, *Knowledge Discovery in Bibliographic Databases*, volume 48 of *Library Trends*. Graduate School of Library and Information Science. University of Illinois at Urbana-Champaign.

Numata, M. (1961). Forest vegetation in the vicinity of choshi. coastal flora and vegetation at choshi, chiba prefecture iv. In *Bulletin of Choshi Marine Laboratory*, volume 3, pages 23–48. Chiba University. in Japanese.

Nunes, K. P. (2009). *Estudos Arqueométricos do Sítio Arqueológico Hatahara*. PhD thesis, Instituto de Pesquisas Energéticas e Nucleares.

Nunes, K. P., Toyota, R. G., Oliveira, P. M. S., Neves, E. G., Soares, E. A. A., and Munita, C. S. (2013). Preliminary compositional evidence of provenance of ceramics from hatahara archaeological site, central amazonia. *Journal of Chemistry*, 2013:1–6.

Obudho, R. A. (1976). Temporal periodicity and locational spacing of periodic and daily markets in Kenya. *Cahiers d'études africaines*, 16(63):553–566.

Ohser, J. (1983). On estimators for the reduced second moment measure of point processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 14(1):63–71.

Oliver, J. R. (2008). The archaeology of agriculture in ancient amazonia. In Silverman and Isbell (2008).

Openshaw, S., Charlton, M., Wymer, C., and Craft, A. (1987). A mark 1 geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems*, 1(4):335–358.

O'Sullivan, D. and Unwin, D. J. (2010). *Geographic information analysis*. John Wiley & Sons, 2nd edition.

Paliou, E. (2008). An autonomous agent approach to the investigation of intra-site movement and visibility: The visual consumption of theran murals from the public spaces of lba akrotiri (thera, greece). In Posluschny, A., Lambers, K., and Herzog, I., editors, *Proceedings of the Annual International conference of Computer applications and quantitative methods in Archaeology, CAA 2007, Berlin, 2-6, April 2007*, pages 328–335.

Palmisano, A. (2013). Zooming patterns among the scales: a statistics technique to detect spatial patterns among settlements. In *Archaeology in the Digital Era: Proceedings of the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology*, page 348.

Pappu, R. S. (1988). Site catchment analysis. In Dhavalikar, M. K., Sankalia, H. D., and Ansari, Z., editors, *Excavations at Inamgaon*, volume 1, pages 107–120. Deccan College.

Pappu, R. S. and Shinde, V. S. (1990). Site catchment analysis of the deccan chalcolothic in the central tapi basin. *Bulletin of the Deccan College Research Institute*, 49:317–338.

Pardi, M. L. F. (2002). *Gestão de patrimônio arqueológico, documentação e política de preservação*. PhD thesis, Universidade Católica de Goiás.

Parker, S. (1985). Predictive modelling of site settlement systems using multivariate logistics. In Carr, C., editor, *For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Techniques, and Theory*, pages 173–207. Westport Publishers St. Louis, Kansas City.

Parsons, J. R. (1972). Archaeological settlement patterns. *Annual Review of Anthropology*, 1:127–150.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302):157–175.

Peebles, C. S. and Kus, S. M. (1977). Some archaeological correlates of ranked societies. *American Antiquity*, 42(3):421–448.

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4):231–259.

Phillips, S. J. and Dudík, M. (2008). Modeling of species distributions with maxent: new extensions and a comprehensive evaluation. *Ecography*, 31(2):161–175.

Phillips, S. J., Dudík, M., and Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. In *Proceedings of the Twenty-first International Conference on Machine Learning*, ICML '04, pages 655–662, New York, NY, USA. ACM.

Pickard, D. K. (1982). Isolated nearest neighbors. *Journal of Applied Probability*, 19(2):444–449.

Pielou, E. C. (1959). The use of point-to-plant distances in the study of the pattern of plant populations. *Journal of Ecology*, 47(3):607–613.

Pielou, E. C. (1969). *An Introduction to Mathematical Ecology*. John Wiley & Sons, New York.

Pielou, E. C. (1977). *Mathematical Ecology*. Wiley-Interscience publication. John Wiley & Sons.

Pizziolo, G. (2015). Predicting prehistory: Themes, problems, perspectives. In Pizziolo, G. and Sarti, L., editors, *Predicting Prehistory - Predictive Models and Field Research Methods for Detecting Prehistoric Context*, number 11 in Millenni Studi di archeologia preistorica, pages 11–22. Museo e Instituto Fiorentino di Preistoria "Paolo Graziosi", Florence.

Plog, F. T. (1981). Sarg: The computer in a cooperative effort. In *Computer data bank applications in archaeology*, pages 46–56.

Plog, F. T. (1983). Study area 1: Kaibab national forest. In *Theory and model building : refining survey strategies for locating prehistoric heritage resources*, pages 63–66. U.S. Forest Service, Southwestern Region.

Porro, A. (1994). Social organization and political power in the amazon floodplain: the ethnohistorical sources. In Roosevelt, A. C., editor, *Amazonian Indians from Prehistory to the Present: Anthropological Perspectives*, pages 79–94. University of Arizona Press.

Powell, V. and Lehe, L. (2015). Principal component analysis.

Prance, G. T. and Lovejoy, T. E., editors (1985). *Key Environments - Amazonia*. Oxford Pergamon Press.

PRONAPA (1966). Terminologia arqueológica brasileira para a cerâmica. *Manual de Arqueologia*, 1:73.

PRONAPA (1970). Brazilian archaeology in 1968: An interim report on the national program of archaeological research. *American Antiquity*, 35(1):1–23.

Prothero, R. M. (1957). Land use at soba, zaria province, northern nigeria. *Economic Geography*, Vol. 33:72–86. R. Mansell Prothero.

Punj, G. and Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20(2):134–148.

Py-Daniel, A. R. (2009). *Arqueologia da Morte no sítio Hatahara durante a fase paredão*. PhD thesis, Universidade de São Paulo.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850.

Ratnaparkhi, A. (1997). A simple introduction to maximum entropy models for natural language processing. Technical report, Institute for Research in Cognitive Science. IRCS Report 97–08.

Ratnaparkhi, A. (1998). *Maximum Entropy Models for Natural Language Ambiguity Resolution*. PhD thesis, University of Pennsylvania, Philadelphia, PA, USA.

Rebellato, L. (2007). *Interpretando a variabilidade cerâmica e as assinaturas químicas e fÃsicas do solo no SÃtio Arqueológico Hatahara*. PhD thesis, Universidade de São Paulo.

Rebellato, L., Woods, W. I., and Neves, E. G. (2009). Pre-columbian settlement dynamics in the central amazon. In Woods et al. (2009), pages 15–31.

Rencher, A. C. and Christensen, W. F. (2012). *Methods of Multivariate Analysis*. Wiley, 3rd edition.

Renfrew, C. and Bahn, P. (2005). *Archaeology: The Key Concepts*. Routledge key guides. Routledge.

Renfrew, C. and Bahn, P. (2012). *Archaeology: Theories, Methods, and Practice*. Thames & Hudson, 6th edition.

Renfrew, C. and Level, E. V. (1979). Exploring dominance: Predicting polities from centers. In Renfrew, C. and Cooke, K. L., editors, *Transformations: Mathematical Approaches to Culture Change*, pages 145–166. Academic Press, New York.

Rinzivillo, S., Turini, F., Bogorny, V., Körner, C., Kuijpers, B., and May, M. (2008). Knowledge discovery from geographical data. In Giannotti, F. and Pedreschi, D., editors, *Mobility, Data Mining and Privacy*, pages 243–265. Springer, Berlin, Heidelberg.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13(2):255–266.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2):172–212.

Ripley, B. D. (1981). Spatial statistics. *Wiley Series in Probability and Statistics*.

Ripley, B. D. (1987). *Stochastic Simulation*. Wiley Series in Probability and Statistics. Wiley & Sons, Inc.

Ripley, B. D. (1988). *Statistical inference for spatial processes*. Cambridge university press.

Rivers, R., Knappett, C., and Evans, T. (2011). Network models and archaeological spaces. *Computational Approaches to Archaeological Spaces*.

Robertshaw, P. (1994). Archaeological survey, ceramic analysis, and state formation in western uganda. *African Archaeological Review*, 12(1):105–131.

Robinson, D., Murray-Rust, D., Rieser, V., Milicic, V., and Rounsevell, M. (2012). Modelling the impacts of land system dynamics on human well-being: Using an agent-based approach to cope with data limitations in koper, slovenia. *Computers, Environment and Urban Systems*, 36(2):164–176. Special Issue: Geoinformatics 2010.

Rodrigues, J. H., Wirth, J. D., and Whigham, T. L. (1984). An interview with josé honório rodrigues. *The Hispanic American Historical Review*, 64(2):217–232.

Rogers, S. R. and Curdy, P. (2015). Least cost path analysis for predicting glacial archaeological site potential: Scale parameter investigations. In Pizziolo, G. and Sarti, L., editors, *Predicting Prehistory - Predictive Models and Field Research*

Methods for Detecting Prehistoric Context, number 11 in Millenni Studi di archeologia preistorica, pages 49–64. Museo e Instituto Fiorentino di Preistoria "Paolo Graziosi", Florence.

Rokach, L. and Maimon, O. Z. (2005). Clustering methods. In Maimon, O. Z. and Rokach, L., editors, *Data Mining and Knowledge Discovery Handbook*, pages 321–352. Springer.

Rönz, B. and Strohe, H. G., editors (1994). *Lexikon Statistik*. Gabler Wirtschaft. Gabler, Wiesbaden.

Roosevelt, A. C. (1980). *Parmana: Prehistoric Maize and Manioc Subsistence Along the Amazon and Orinoco*. Studies in archaeology. Academic Press.

Roosevelt, A. C. (1991a). Determinismo ecológico na interpretação do desenvolvimento indígena da amazônia. In Neves, W. A., editor, *Origens, adaptações e diversidade biológica do homen nativo da Amazônia*. Museu Paraense Emílio Goeldi. Belém.

Roosevelt, A. C. (1991b). *Moundbuilders of the Amazon. Geophysical Archaeology on Marajo Island, Brazil*. Academic Press.

Roosevelt, A. C. (1994). Amazonian anthropology: strategy for a new synthesis. In Roosevelt, A. C., editor, *Amazonian Indians from prehistory to the present: anthropological perspectives*, pages 1–29. University of Arizona Press, Tucson.

Roosevelt, A. C. (1995). Early pottery in the amazon. twenty years of scholarly obscurity. In Barnett, W. K. and Hoopes, J. W., editors, *The emergence of pottery : technology and innovation in ancient societies*, pages 115–131. Smithsonian Institution Press.

Roosevelt, A. C. (1997). *The Excavations at Corozal, Venezuela: Stratigraphy and Ceramic Seriation*. Number 83. Yale University Publications in Anthropology, New Haven.

Roosevelt, A. C., Lima da Costa, M., Lopes Machado, C., Michab, M., Mercier, N., Valladas, H., Feathers, J., Barnett, W., Imazio da Silveira, M., Henderson, A., and et al. (1996). Paleoindian cave dwellers in the amazon: The peopling of the americas. *Science*, 272(5260):373–384.

Roper, D. C. (1974). *The distribution of middle woodland sites within the environment of the Lower Sangamon River, Illinois*. Number 30 in Reports of investigations (Illinois State Museum). Illinois State Museum, Springfiled.

Roper, D. C. (1976). A trend-surface analysis of central plains radiocarbon dates. *American Antiquity*, 41(2):181–189.

Roper, D. C. (1979a). *Archaeological survey and settlement pattern models in central Illinois*, volume 16. Kent State University Press.

Roper, D. C. (1979b). The method and theory of site catchment analysis: A review. *Advances in Archaeological Method and Theory*, 2:119–140.

Scarry, J. F. and Payne, C. (1986). Mississippian polities in the fort walton area: A model generated from the renfrew-level xtent algorithm. *Southeastern Archaeology*, 5(2):79–90.

Schaan, D. P. (2001). Os dados inéditos do projeto marajó (1962-1965). *Revista do Museu de Arqueologia e Etnologia*, (11):141–164.

Schaan, D. P. (2004). *The Camutins chiefdom: rise and development of social complexity on Marajó Island, Brazilian Amazon*. PhD thesis, University of Pittsburgh.

Schaan, D. P. (2007). Uma janela para a história pré-colonial da amazônia: olhando além - e apesar - das fases e tradições. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas*, 2(1):77–89.

Schilling, M. F. (1986). Mutual and shared neighbor probabilities: Finite- and infinite-dimensional results. *Advances in Applied Probability*, 18(2):388–405.

Schlotzhauer, S. D. (2007). *Elementary Statistics Using JMP®*. SAS Press. SAS Institute.

Schmidt Dias, A. (1995). Um projeto para a arqueologia brasileira: breve histórico da implementação do pronapa. *Revista do CEPA, Santa Cruz do Sul*, 19(22):25–40.

Scholtz, S. C. (1981). Location choice models in sparta. In Lafferty III, R. H., Otinger, J. L., Scholtz, S. C., Limp, W. F., Watkins, B., and Jones, R. D., editors, *Settlement Predictions in Sparta: A Locational Analysis and Cultural Resource Assessment in the Uplands of Calhoun County, Arkansas*, volume 14 of *Archaeological Survey Research Series*, pages 207–222. Arkansas.

Schönwiese, C.-D. (1992). *Praktische Statistik für Meteorologen und Geowissenschaftler*. Schweizerbart Science Publishers, Stuttgart, Germany, 2nd edition.

Schuurman, N., Fiedler, R. S., Grzybowski, S. C. W., and Grund, D. (2006). Defining rational hospital catchments for non-urban areas based on travel-time. *International Journal of Health Geographics*, 5(1):1–11.

Sebastian, L. and Judge, W. J. (1988). Predicting the past: Correlation, explanation, and the use of archaeological models. In Judge, J. W. and Sebastian, L., editors, *Quantifying the Present and Predicting the Past: Theory, Method and Application of Archaeological Predictive Modelling*, pages 1–18. U.S. Department of the Interior, Bureau of Land Management, Denver.

Şenbabaoğlu, Y., Michailidis, G., and Li, J. Z. (2014). Critical limitations of consensus clustering in class discovery. *Scientific Reports*, 4(6207):1–13.

Sevenant, M. and Antrop, M. (2007). Settlement models, land use and visibility in rural landscapes: Two case studies in greece. *Landscape and Urban Planning*, 80(4):362–374.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Silva, F. and Steele, J. (2014). New methods for reconstructing geographical effects on dispersal rates and routes from large-scale radiocarbon databases. *Journal of Archaeological Science*, 52:609–620.

Silverman, H. and Isbell, W., editors (2008). *The Handbook of South American Archaeology*. Springer, New York.

Simões, M. F. (1972). Índice das fases arqueológicas brasileiras: 1950-1971. *Publicações Avulsas do Museu Paraense Emílio Goeldi. Belém*, 18:1–75.

Simões, M. F. (1981). Coletores-pescadores ceramistas do litoral do salgado (pará). *Nota Preliminar. Boletim do Museu Paraense Emílio Goeldi, Nova Série, Antropologia, Belém*, 78:1–26.

Simões, M. F. (1983). Pesquisa e cadastro de sítios arqueológicos na amazônia legal brasileira 1978-1982. *Publicações Avulsas do Museu Paraense Emílio Goeldi. Belém*.

Simões, M. F. and Araújo-Costa, F. d. (1978). Áreas da amazônia legal brasileira para pesquisa e cadastro de sítios arqueológicos. *Publicações Avulsas do Museu Paraense Emílio Goeldi. Belém*.

Simões, M. F. and Corrêa, C. G. (1987). Pesquisas arqueológicas no baixo uatumã-jatapu (amazonas). *Revista de Arqueologia*, 4(1):29–48.

Simões, M. F. and de Araujo-Costa, F. (1987). Pesquisas arqueológicas no baixo rio tocantins (pará). *Revista de arqueologia*, 4(1):11–27.

Simões, M. F. and Kalkmann, A. L. M. (1987). Pesquisas arqueológicas no médio rio negro (am). *Revista de Arqueologia*, 4(1):83–116.

Simões, M. F. and Lopes, D. F. (1987). Pesquisas arqueológicas no baixo/médio rio madeira (am). *Revista de Arqueologia*, 4(1):117–134.

Simões, M. F. and Machado, A. L. (1987). Pesquisas arqueológicas no lago de silves (am). *Revista de Arqueologia*, 4(1):49–82.

Sioli, H. (1983). *Amazonien. Grundlagen der Ökologie des grössten tropischen Waldlandes.* Wissenschafltiche Verlagsgesellschaft mbH, Stuttgart.

Smith, N. J. H. (1980). Anthrosols and human carrying capacity in amazonia. *Annals of the Association of American Geographers*, 70(4):553–556.

Smith, R. H. T. (1971). West african marketplaces: Temporal periodicity and locational spacing. In Meillassoux, C., editor, *Development of Indigenous Trade and Markets in West Africa*, pages 319–346. Oxford University Press.

Soetens, S., Sarris, A., Vansteenhuyse, K., and Topouzi, S. (2003). Gis variations on a cretan theme: Minoan peak sanctuaries. In *Metron: Measuring the Aegean Bronze Age. Proceedings of the 9th International Aegean Conference, New Haven, Yale University, 18-21 April 2002 (Aegaeum)*, pages 483–488.

Sollars, L. (2004). Settlement in the prehistoric mediterranean. In Blake, E. and Knapp, A. B., editors, *The Archaeology of Mediterranean Prehistory*, number 6 in Blackwell Studies in Global Archaeology, pages 252–269. Blackwell Publishing Ltd.

Stančič, Z. and Kvamme, K. L. (1999). Settlement pattern modelling through boolean overlays of social and environmental variables. In Barceló, J. A., Briz, I., and Vila, A., editors, *New Techniques for Old Times. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998*, number 757 in BAR International Series, pages 231–237. Archaeopress, Oxford.

Stanislawski, M. B. (1973). Review: Archaeology as anthropology: A case study by william a. longacre. *American Antiquity*, 38(1):117–122.

Steponaitis, V. P. (1981). Settlement hierarchies and political complexity in nonmarket societies: The formative period of the valley of mexico. *American Anthropologist*, 83(2):320–363.

Steward, J. H. (1938). *Basin-plateau aboriginal sociopolitical groups*, volume 120. Smithsonian Institution Bureau of American Ethnology Bulletin.

Steward, J. H., editor (1940-1947). *Handbook of South American Indians*, volume 1 - 6. Smithsonian Institution. 1940-1947.

Stiteler, W. M. and Patil, G. P. (1971). Variance to mean ratio and morisita's index as measures of spatial patterns in ecological populations. *Statistical Ecology*, 1:423–459.

Strauss, D. J. (1975). A model for clustering. *Biometrika*, 62(2):467–475.

Struever, S. and Houart, G. L. (1972). Analysis of the hopewell interaction sphere. In *Social exchange and interaction*, volume 46 of *Anthropological papers of the Museum of Anthropology*, pages 47–79. University of Michigan.

Styles, B. W. (1981). *Faunal exploitation and resource selection: Early Late Woodland subsistence in the lower Illinois Valley*. Northwestern University Archeological Program.

Su, S., Zhang, Q., Zhang, Z., Zhi, J., and Wu, J. (2011). Rural settlement expansion and paddy soil loss across an ex-urbanizing watershed in eastern coastal china during market transition. *Regional Environmental Change*, 11(3):651–662.

Svenning, J.-C. and Skov, F. (2004). Limited filling of the potential range in european tree species. *Ecology Letters*, 7(7):565–573.

Tanser, F., Gijsbertsen, B., and Herbst, K. (2006). Modelling and understanding primary health care accessibility and utilization in rural south africa: An exploration using a geographical information system. *Social Science & Medicine*, 63(3):691–705.

Taylor, W. W. (1948). A study of archaeology. *American Anthropological Association*, 69.

Tian, G., Qiao, Z., and Zhang, Y. (2012). The investigation of relationship between rural settlement density, size, spatial distribution and its geophysical parameters of china using landsat {TM} images. *Ecological Modelling*, 231:25–36.

Tiffany, J. A. and Abbott, L. R. (1982). Site-catchment analysis: Applications to iowa archaeology. *Journal of Field Archaeology*, 9(3):313–322.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240.

Tobler, W. R. (1993). Three presentations on geographical analysis and modeling: Non-isotropic geographic modeling, speculations on the geometry of geography, global spatial analysis. Technical Report 93-1, National Center for Geographic Information and Analysis.

Tomlin, C. D. (1991). Cartographic modelling. In D J Maguire, M F Goodchild, D. W. R., editor, *Geographical Information Systems: Principles and Applications*, volume Vol. 1, chapter 23, pages 361–374. Longman Scientific and Technical, Harlow, UK.

Tryfos, P. (1998). *Methods for Business Analysis and Forecasting: Text and Cases*. Wiley.

Ueberschär, N. (2013). Gis-methoden zur schätzung von einzugsgebieten für gesundheitszentren in ruanda. In Strobl, J., Blaschke, T., Griesebner, G., and Zagel, B., editors, *AGIT 2013 - Symposium und Fachmesse Angewandte Geoinformatik*, pages 338–343.

Ullah, I. I. (2011). A gis method for assessing the zone of human-environmental impact around archaeological sites: a test case from the late neolithic of wadi ziqlâb, jordan. *Journal of Archaeological Science*, 38(3):623–632.

Unwin, D. J. (1975). *An Introduction to Trend Surface Analysis*. Number 5 in Concepts and Techniques in Modern Geography (CATMOG). Geo Abstracts, University of East Anglia, Norwich.

Van Leusen, M. (1999). Viewshed and cost surface analysis using gis (cartographic modelling in a cell-based gis ii). In Barceló, J. A., Briz, I., and Vila, A., editors, *New Techniques for Old Times. CAA98. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 26th Conference, Barcelona, March 1998*, volume 757 of *BAR International Series*, pages 215–224, Oxford. Archaeopress.

Van Leusen, M., Deeben, J., Hallewas, D., Zoetbrood, P., Kamermans, H., and Verhagen, P. (2005). A baseline for predictive modelling in the netherlands. In *Predictive modelling for archaeological heritage management: A research agenda*, pages 25–93. Rijksdienst voor het Oudheidkundig Bodemonderzoek.

Van Leusen, P. M., Pizziolo, G., and Sarti, L., editors (2011). *Hidden Landscapes of Mediterranean Europe. Cultural and methodological biases in pre- and protohistoric landscape studies - Proceedings of the international meeting Siena, Italy, May 25-27, 2007*.

Verhagen, P., editor (2007). *Case Studies in Archaeological Predictive Modelling*. Number 14 in Archaeological Studies Leiden University Press Series. Leiden University Press.

Verhagen, P., Kamermans, H., van Leusen, M., Deeben, J., Hallewas, D., and Zoetbrood, P. (2007). First thoughts on the incorporation of cultural variables into predictive modelling. In Verhagen, P., editor, *Case Studies in Archaeological Predictive Modelling*, number 14 in Archaeological Studies Leiden University Press Series, chapter 10, pages 203–210. Leiden University Press.

Verhagen, P., Nuninger, L., Tourneux, F.-P., Bertoncello, F., and Jeneson, K. (2013). Introducing the human factor in predictive modelling: a work in progress. In *Archaeology in the Digital Era, Proceedings of CAA Conference, 40th Annual Meeting, Southampton, UK*, pages 379–388.

Verhagen, P. and Whitley, T. G. (2012). Integrating archaeological theory and predictive modeling: a live report from the scene. *Journal of Archaeological Method and Theory*, 19(1):49–100.

Vita-Finzi, C. and Higgs, E. (1970). Prehistoric economy in the mount carmel area of palestine: Site catchment analysis. In *Proceedings of the Prehistoric Society*, volume 36, pages 1–37.

Viveiros de Castro, E. (1996). Images of nature and society in amazonian ethnology. *Annual Review of Anthropology*, 25:179–200.

von Groote-Bidlingmaier, C., Hilbert, K., Schwer, J., and Timpf, S. (2015). Interactive webgis for archaeological settlement pattern analysis - a requirement analysis. In *ICC 2015*.

von Groote-Bidlingmaier, C., Sabine, T., and Hilbert, K. (2014a). Erstellung von potenziellen einflussgebieten auf der basis von archäologischen ausgrabungsorten - am beispiel prähistorischer kulturen im amazonasgebiet. In Strobl, J., Blaschke, J., Griesebner, G., and Zagel, B., editors, *Angewandte Geoinformatik 2014 - Beiträge zum 26. AGIT-Symposium Salzburg*, pages 77–86.

von Groote-Bidlingmaier, C., Sabine, T., and Hilbert, K. (2014b). Estimating moving regions out of point data - from excavation sites in the amazon region to areas of influence of prehistoric cultures. In Huerta, J., Schade, S., and Granell, C., editors, *Connecting a Digital Europe through Location and Place. Proceedings of the AGILE 2014 International Conference on Geographic Information Science, Castellón, June, 3-6, 2014*.

von Thünen, J. (1875). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. 3rd edition.

Wagner, S. and Wagner, D. (2007). Comparing clusterings - an overview. Technical Report 2006-04, Universität Karlsruhe (TH).

Waibel, L. (1958). *Capítulos de geografia tropical e do Brasil*. Coleção Ibgeana, Rio de Janeiro.

Wang, S.-C. (2003). Artificial neural network. In *Interdisciplinary Computing in Java Programming*, volume 743 of *The Springer International Series in Engineering and Computer Science*, pages 81–100. Springer US.

Wansleeben, M. and Verhart, L. (1997). Geographical information systems - methodological progress and theoretical decline? *Archaeological Dialogues*, 4(01):53–70.

Warren, R. E. (1990). Predictive modelling of archaeological site location: a case study in the midwest. In Allen, K. M., Green, S. W., and Zubrow, E. B. W., editors, *Interpreting space: GIS and archaeology*, chapter 18, pages 201–215. Taylor & Francis.

Washburn, D. K. (1974). Nearest neighbor analysis of pueblo i-iii settlement patterns along the rio puerco of the east, new mexico. *American Antiquity*, 39(2):315–335.

Wei, W., Shi, P., Zhou, J., Feng, H., Wang, X., and Wang, X. (2013). Environmental suitability evaluation for human settlements in an arid inland river basin: A case study of the shiyang river basin. *Journal of Geographical Sciences*, 23(2):331–343.

Wheatley, D. (2004). Making space for an archaeology of place. *Internet Archaeology*, (15).

Wheatley, D. and Gillings, M. (2002). *Spatial Technology and Archaeology: The Archaeological Applications of GIS*. Routledge, London, New York.

Whitehead, N. L. (2010). Amazonia. In McNeill, W. H., editor, *Berkshire Encyclopedia of World History*, pages 97–100. Berkshire Publishing Group, 2nd edition.

Whitley, T. G. (2005). A brief outline of causality-based cognitive archaeological probabilistic modeling. In Van Leusen, P. M. and Kamermans, H., editors, *Predictive Modelling for Archaeological Heritage Management: A Research Agenda*, pages 123–137. National Service for Archaeological Heritage.

Whitley, T. G. and Burns, G. (2008). Conditional gis surfaces and their potential for archaeological predictive modelling. In *CAA2007 - Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, Germany, April 2-6, 2007*, pages 292–298.

Wilkerson, M. D. and Hayes, D. N. (2010). Consensusclusterplus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*, 26(12):1572–1573.

Willey, G. and Phillips, P. (1958). *Method and Theory in American Archaeology*. The University of Chicago Press.

Willey, G. R. (1945). Horizon styles and pottery traditions in peruvian archaeology. *American Antiquity*, 11(1):49–56.

Willey, G. R. (1953). *Prehistoric settlement patterns in the Virú Valley, Peru*, volume 155. Smithsonian Institution.

Willey, G. R. (1962). The early great styles and the rise of the pre-columbian civilizations. *American Anthropologist*, 64(1):1–14.

Willey, G. R. and Sabloff, J. (1993). *A History of American Archaeology*, volume 3rd. Freeman, W.H.

Williams, L., Thomas, D. H., and Bettinger, R. (1973). Notions to numbers: Great basin settlements as polythetic sets. *Research and Theory in Current Archaeology*, pages 215–237.

WinklerPrins, A. and Aldrich, S. (2010). Locating amazon dark earths: Creating an interactive gis of known locations. *Journal of Latin American Geography*, 9(3):33–50.

Wood, J. j. (1971). Fitting discrete probability distributions to prehistoric settlement patterns. In Gumerman, G. J., editor, *The Distribution of Prehistoric Population Aggregates*, number 1 in Prescott College Anthropological Reports.

Woods, W. I. and Denevan, W. M. (2009). Amazonian dark earths: The first century of reports. In Woods et al. (2009), pages 1–14.

Woods, W. I. and Glaser, B. (2004). Towards an understanding of amazonian dark earths. In Glaser, B. and Woods, W. I., editors, *Amazonian Dark Earths: Explorations in Space and Time*, pages 1–8. Springer, Berlin Heidelberg.

Woods, W. I. and McCann, J. M. (1999). The anthropogenic origin and persistence of amazonian dark earths. In *Yearbook. Conference of Latin Americanist Geographers*, volume 25, pages 7–14. University of Texas Press.

Woods, W. I., Teixeira, W. G., Lehmann, J., Steiner, C., WinklerPrins, A., and Rebellato, L., editors (2009). *Amazonian Dark Earths: Wim Sombroekś Vision*. Springer.

Wrigley, N. (1977). Probability surface mapping: A new approach to trend surface mapping. *Transactions of the Institute of British Geographers*, 2(2):129–140.

Xiao, R., Su, S., Wang, J., Zhang, Z., Jiang, D., and Wu, J. (2013). Local spatial modeling of paddy soil landscape patterns in response to urbanization across the urban agglomeration around hangzhou bay, china. *Applied Geography*, 39:158–171.

Yuan, M., Buttenfield, B., Gahegan, M., and Miller, H. (2004). Geospatial data mining and knowledge discovery. In McMaster, R. B. and Usery, E. L., editors, *A Research Agenda for Geographic Information Science*, chapter 14, pages 365–388. CRC Press.

Zhang, Z., Su, S., Xiao, R., Jiang, D., and Wu, J. (2013). Identifying determinants of urban growth from a multi-scale perspective: A case study of the urban agglomeration around hangzhou bay, china. *Applied Geography*, 45:193–202.

Zhang, Z., Xiao, R., Shortridge, A., and Wu, J. (2014). Spatial point pattern analysis of human settlements and geographical associations in eastern coastal china – a case study. *International Journal of Environmental Research and Public Health*, 11(3):2818–2833.

Zubrow, E. B. and Harbaugh, J. W. (1978). Archaeological prospecting: kriging and simulation. In Hodder, I., editor, *Simulation Studies in Archaeology*, pages 109–122. Cambridge University Press, Cambridge.

A IPHAN Database Output

Ministério da Cultura
Sistema Nacional de Informações Culturais - SNIC

Cadastro Nacional de Sítios Arqueológicos
CNSA / SGPA*

IPHAN
Centro Nacional de Arqueologia - CNA

- CNSA AM00002 -

Nome do sítio: AM-UR-9: Nova Estrela
Outras designações e siglas: AM-UR-9 **CNSA**: AM00002
Município: Itapiranga **UF**: AM
Descrição sumária do sítio: Sítio-Habitação localizado entre a margem direita do rio Uatumã e a esquerda do igarapé Caiuazinho.
Sítios relacionados:

Comprimento: 200m **Largura**: 90m **Altura máxima**: 39m (a partir do nível do solo)
Área: $0m^2$ **Medição** ○ **Estimada** ○ **Passo** ○ **Mapa** ◉ **Instrumento**
Unidade geomorfológica: Andulado
Compartimento topográfico: Topo de Colina
Altitude: 0**m**(com relação ao nível do mar)
Água mais próxima: Igarapé Caiuazinho
Distância: 0**m**
Rio: Uatumã
Bacia:

Vegetação atual
☐ **Floresta ombrófila** ☐ **Savana (cerrado)**
☐ **Floresta estacional** ☐ **Savana-estépica (Caatinga)**
☐ **Campinarana** ☐ **Estepe**
☐ **Capoeira** **Outra**: Mata

Uso atual do terreno
☐ **Atividade urbana** ☐ **Pasto**
☐ **Via pública** ☑ **Plantio**
☐ **Estrutura de fazenda** ☐ **Área devoluta**
Outro:

Propriedade da terra ☐ **Área pública** ☑ **Área privada** ☐ **Área militar** ☐ **Área indígena**
Outra:
Proteção legal ☐ **Unid. de conservação ambiental**
Em área tombada ☐ **Municipal** ☐ **Estadual** ☐ **Federal** ☐ **Patrim. da humanidade**

Categoria
◉ **Unicomponencial** ☑ **Pré-colonial**
○ **Multicomponencial** ☐ **De contato**
☐ **Histórico**

Tipo de sítios:
Forma: Elipsoidal
Tipo de solo:

Estratigrafia:
Contexto de deposição ☑ **Em superfície** ☐ **Em profundidade**
Exposição ◉ **Céu aberto** ○ **Abrigo sob rocha** ○ **Gruta** ○ **Submerso**
○ **Outra**:

* Em atendimento ao determinado na Lei nº 3.924 de 26 de julho de 1961, que dispõe sobre os monumentos arqueológicos e pré-históricos.

Página 1 de 3

Cadastro Nacional de Sítios Arqueológicos*

Sist. Nac. de Patrimônio Cultural - SNPC | - CNSA AM00002 - | Centro Nacional de Arqueologia - CNA

Estrutura

- [x] **Área de refugo**
- [ ] **Canais tipo trincheiras, valetas**
- [ ] **De lascamento**
- [ ] **Círculos de pedra**
- [ ] **De Combustão** (fogueira, forno, fogão)
- [ ] **Estacas, buracos de Fossas**
- [ ] **Funerárias**
- [ ] **Fossas**
- [ ] **Vestígios de edificações**
- [ ] **Muros de terra, linhas de argila**
- [ ] **Vestígios de mineração**
- [ ] **Palafitas**
- [ ] **Alinhamento de pedras**
- [ ] **Paliçadas**
- [x] **Manchas pretas**
- [ ] **Concentrações cerâmica - quant.**:

Outras:

Artefatos

- [ ] **Lítico lascado**
- [x] **Cerâmico**
- [ ] **Lítico polido**
- [ ] **Sobre concha**
- [ ] **Sobre material orgânico**

Outros vestígios líticos:

Material histórico:
Outros vestígios orgânicos:
Outros vestígios inorgânicos:

Arte rupestre: [ ] **Pintura**: [ ] **Gravura**: [x] **Ausente**:

FILIAÇÃO CULTURAL

Artefatos líticos:
Tradições:
Fases:
Complementos:
Outras atribuições:

Artefatos cerâmicos:
Tradições: Regional Saracá
Fases: Uatumã
Complementos:
Outras atribuições:

Artefatos rupestre:
Tradições:
Estilos:
Complementos:
Outras atribuições:

Datações Absolutas:
Datações Relativas:
Grau de integridade ( ) **mais de 25%** ( ) **entre 25 e 75%** (x) **menos de 25%**
Fatores de destruição
- [ ] **Erosão eólica**
- [ ] **Erosão fluvial**
- [ ] **Vandalismo**
- [ ] **Erosão pluvial**
- [ ] **Atividades agrícolas**
- [ ] **Construção de estrada**
- [ ] **Construção de moradias**

Outros fatores naturais:
Outros fatores antrópicos: Atividade agrícola e as queimadas
Possibilidades de destruição: Sim
Medidas para preservação:
Relevância do sítio ( ) **Alta** ( ) **Média** (x) **Baixa**

Atividades desenvolvidas no local
- [x] **Registro**
- [x] **Coleta de superfície**
- [ ] **Sondagem ou Corte estratigráfico**
- [ ] **Escavação de grande superfície**
- [ ] **Levantamento de grafismo rupestre**

Nome do responsável pelo registro: Mário F.Simões, Conceição Corrêa e Ana L. Maroja
Data do registro: 22/11/1979 **Ano do registro**:

* Em atendimento ao determinado na Lei nº 3.924 de 26 de julho de 1961, que dispõe sobre os monumentos arqueológicos e pré-históricos.

Cadastro Nacional de Sítios Arqueológicos*

Sist. Nac. de Patrimônio Cultural - SNPC | **- CNSA AM00002 -** | Centro Nacional de Arqueologia - CNA

Nome do projeto:

Documentação produzida (quantidade)

Mapa com sítio plotado:	10	**Foto preto e branco**:	20
Croqui:	0	**Reprografia de imagem**:	0
Planta baixa do sítio:	0	**Imagem de satélite**:	0
Planta baixa dos locais afetados:	0	**Cópia total de arte rupestre**:	0
Planta baixa de estruturas:	0	**Cópia parcial de arte rupestre**:	0
Perfil estratigráfico:	0	**Ilustração do material**:	0
Perfil topográfico:	0	**Caderneta de campo**:	0
Foto aérea:	0	**Video / Filme**:	0
Foto colorida:	0	**Outra**:	0

Bibliografia

Simões, Mário F. 1979. Pesquisas Arqueológicas nos rio Urubu, Uatumã e Jatapu. Relatório Preliminar. Belém-Pa, Museu Paraense Emílio Goeldi, 101 p. il.

Simões, Mário F. 1978-1982. Pesquisa e Cadastro de Sítios Arqueológicos na Amazônia Legal Brasileira, Publicações Avulsas do Museu Pa. Emílio Goeldi. Belém-Pa, 100 p.

Simões, Mário F. e Corrêa, Conceição G. 1987. Pesquisas Arqueológicas no Baixo Uatumã-Jatapu (AM). In: Revista de Arqueologia, Belém, Vol. 4, nº 1, p. 29-48. il.

Responsável pelo preenchimento da ficha: Ana Lucia Machado

Data: 09/07/1997

Localização dos dados: Acervo do Museu Emílio Goeldi

Atualizações:

Assinatura

* Em atendimento ao determinado na Lei nº 3.924 de 26 de julho de 1961, que dispõe sobre os monumentos arqueológicos e pré-históricos.

Figure A.01: Output of the IPHAN database for the excavation site AM-UR-9

B Experimental Formulation of Four Horizon Styles

	Zoned-HachureHorizon Style				Incised-Rim Horizon Style						Polychrome Horizon Style						Incised-and-Punctate Horzion Style					
	Tutishcainyo	Yasuni	Jauarí	Ananatuba	Mangueiras	Boím	Manacapurú	Nericagua	Cotua	Los Caros	Napo	Caarí	Araçá	Guarita	Marajoara	Aristé	Mabaruma	Aruquín	Itacoatiara	Santarém	Konduri	Mazagão
Decoration:																						
Incised on rim top	x	?			R		x	x	x	x							x	x	x		x	R
Incised-and-punctate	x	x								R							x	x	x	x	x	x
Incised-and-modeled																	x		x	x		
Incised on parallel lines																	x	x	x	x	x	x
Adornos								R						R	R		x	x	x	x	x	R
Red slip		x		x	x	x	x	x	x	R	x	x	x		x		x	x		x	R	
White slip							x		x		x	x	x	x	x	x			x	x		
Polychrome painting											x	x	x	x	x	x			x	x		
Excised											x		x		x			x	R			
Incised and retouched											x			x	x							
Grooved												x	x	x								
Zone hachure	x	x	x	x																		
Zone red painting	x	x		x																		
Rim form:																						
Interiorly thickened			x		x	x	x	x	x	x								x	x		x	
Everted	x	x							?	?								x	x			
Exteriorly thickend, square												x	R	x	x							
Cambered											x	x	x			x						

Chart showing the occurence of elements of pottery decoration and rim form in complexes representing the four horizon styles. The diagnostic elements of each style are enclosed in boxes.
R signifies that the trait is present but rare; ? That it is probably present.

Table B.01: The experimental formulation of four horizon styles based on 22 pottery complexes (adapted from Meggers and Evans 1961)

C Chronological Sequence of Horizon Styles in the Tropical Forest Area

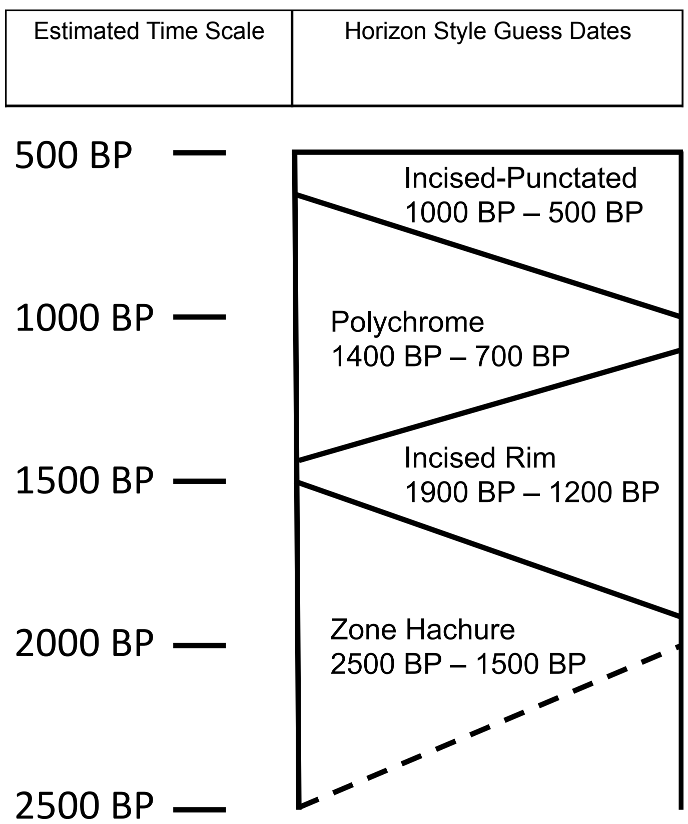


Figure C.01: Chronological sequence of horizon styles in the Tropical Forest Area. Dividing lines are drawn diagonally to suggest the possibility of time lag because of size of the geographical area over which each horizon style spreads. Dates are approximations that may be subject to revision when the absolute chronology becomes better known (adapted from Meggers and Evans 1961)

D Overview of the Known Phases and Traditions in the Amazon in 1972

Estados e Territórios	Fases Cerâmicas				
	Outras Tradições	Tradição Hachurada Zonada	Tradição Borda Incisa	Tradição Policroma	Tradição Incisa Ponteada
Amapá	Aruã Maracá			Aristé	Mazagão
Pará	Acauan Areão (M) Aruã Castália (M) Formiga Itacaiúnas (T) Mina (M) Uruá (M)	Anatuba Jauari	Mangueiras	Marajoara	Konduri
Amazonas	Japurá Santa Luzia		Caiambé Manacapuru Paredão	Guarita Pirapitinga São Joaquim Tefé	Itacoatiara
Mato Grosso (norte)					Diauartum Ipavu

Tradições: M = Mina
T = Tupiguarani

Table D.01: Overview over the known archaeological traditions and phases in the Amazon Basin in 1972 (adapted from Simões 1972)

E Quadrat Methods in Point Processes

No. of flying bombs per square	Expected no. of squares (Poisson)	Actual no. of squares
0	226.74	229
1	211.39	211
2	98.54	93
3	30.62	35
4	7.14	7
5 and over	1.57	1
	576.00	576

Table E.01: Table showing the expected and actual number of flying bombs per square (adapted from Clarke 1946)

Index	Estimator	Reference
I	$\frac{S^2}{\bar{X}}$	Fisher et al. (1922)
ICS	$\frac{S^2}{\bar{X}}-1$	David and Moore (1954)
ICF	$\frac{\bar{X}^2}{S^2-\bar{X}}$	Douglas (1975)
$\dot{X}$	$\bar{X}+\frac{S^2}{\bar{X}}-1$	Lloyd (1967)
IP	$\frac{\dot{X}}{\bar{X}}$	Lloyd (1967)
I_δ	$\frac{n\sum_{i=1}^{n} X_i(X_i-1)}{n\bar{X}(n\bar{X}-1)}$	Morisita (1959)

Table E.02: Table showing optional indices for quadrat count data (adapted from Cressie 1993). X_i is the number of observed objects in a quadrat, $\bar{X}$ is the sample mean of the quadrat counts and S^2 the sample variance.

F Statistical Tables

				α - 1						
df	$\chi_{0.005}$	$\chi_{0.010}$	$\chi_{0.025}$	$\chi_{0.050}$	$\chi_{0.100}$	$\chi_{0.900}$	$\chi_{0.950}$	$\chi_{0.975}$	$\chi_{0.990}$	$\chi_{0.995}$
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Table F.01: Table showing the critical values of the χ^2 distribution $(\alpha - 1)$

	df$_2$									
df$_2$	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18.5	19.0	19.2	19.2	19.3	19.3	19.3	19.4	19.4	19.4
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.36	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88

Table F.02: Table showing the critical values of the F-distribution for an significance level of $\alpha = 5\%$ (Kreyszig 1968)

G Testing the Significance of Regression Functions

BOX 1

The R^2 value is defined as follows:

$$R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total sum of squares}} \tag{G.1}$$

where the residual sum of squares is defined as

$$\sum_{i=1}(y_i - f_i)^2 \tag{G.2}$$

and the total sum of squares as

$$\sum_{i=1}(y_i - \bar{y})^2 \tag{G.3}$$

with:

y_i: examined values of the locations

$\bar{y}$: mean of the values

f_i: used regression function

The result is a value between 0 and 1. (Unwin 1975)

BOX 2

The F value (sometimes referred to as F ratio) is based on the R^2 value but needs the percentage (thus a number between 0 and 100). This can simply be achieved by multiplying the R^2 by 100 which is represented by the variable $\%R^2$. The F value is defined as follows:

$$F = \frac{\%R^2/df_1}{(100 - \%R^2)/df_2} \tag{G.4}$$

with:

df_1: degrees of freedom associated with the surface. They equal the number of coefficients of the used regression function minus 1

df_2: degrees of freedom associated with the residuals. They are equal to the number of observations (total degrees of freedom) minus those degrees of freedom associated with the surface (df_1) minus 1: $df_2 = df_t - df_1 - 1$

(Unwin 1975)

BOX 3

The F value (sometimes referred to as F ratio) is based on the R^2 value but needs the percentage (thus a number between 0 and 100). This can simply be achieved by multiplying the R^2 by 100 which is represented by the variable $\%R^2$. In this case trend surface of order n and order $n+1$ are compared. The modified F value is defined as follows:

$$F = \frac{\%R^2_{extra}/df_3}{(100 - \%R^2_{n+1})/df_{2_{n+1}}} \tag{G.5}$$

with:

$\%R^2_{extra}$: the difference between $\%R^2$ of order $n+1$ and $\%R^2$ of order n, which is the extra $\%R^2$ given by the surface of order $n+1$

df_3: degrees of freedom associated with the added components (3 for a quadratic over a linear function, 4 4 for a cubic over a quadratic, etc.)

$df_{2_{n+1}}$: degrees of freedom associated with the residuals of the order $n+1$ function. They are equal to the number of observations (total degrees of freedom) minus those degrees of freedom associated with the surface (see box 2)

(Unwin 1975)

H Maximum Entropy Modelling

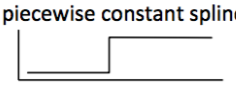
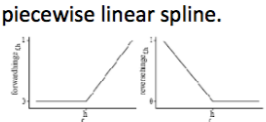
Feature class	Description in relation to environmental variable	Constraint imposed on estimated distribution $\hat{P}$	Ecological interpretation of the constraint
Linear (L)	Variable itself	The mean of variable under $\hat{P}$ should be close to its mean in the sample locations	The mean of the sample indicates average conditions for species presence
Quadratic (Q)	Square of variable	If used with L, variance of variable under $\hat{P}$ is close to its variance in the sample	The variation in that variable in the sample indicates the tolerance of the species for variation from suitable conditions
Product (P)	Product of 2 variables	If used with linear features for the 2 variables, that the covariance of the variables under $\hat{P}$ should be close to its covariance in the sample	The effect of one variable on species presence varies with the value of the other variable – i.e. there are interactions between the variables.
Threshold (T)	A step function that allows a different response below the threshold (the "knot") to that above it. Equivalent to a piecewise constant spline. 	The proportion of $\hat{P}$ that has values of this variable above the knot should be close to that proportion in the sample	Many threshold features can be used on the same variable, with different thresholds. These can add together to model an arbitrary stepped response to the variable.
Hinge (H)	Similar to the threshold feature, but the response above the knot (forward hinge; below left) or below the knot (reverse hinge; below right) is linear with a positive or negative coefficient (slope). Equivalent to a piecewise linear spline. 	The mean of the variable above the knot under $\hat{P}$ should be close to its mean above the knot in the sample locations	A model using only hinge features fits a piecewise linear response. If hinge features are used, linear features are redundant (a linear feature can be created from a hinge, with the knot at one extreme of the feature space).
Category (C)	A binary indicator showing membership in one class of a categorical variable. For a k-class categorical variable there will be k categorical features	The proportion of $\hat{P}$ that has values in this class should be close to that proportion in the sample	

Table H.01: Details about the features in MaxEnt (Elith et al. 2011)

H.1 Konduri

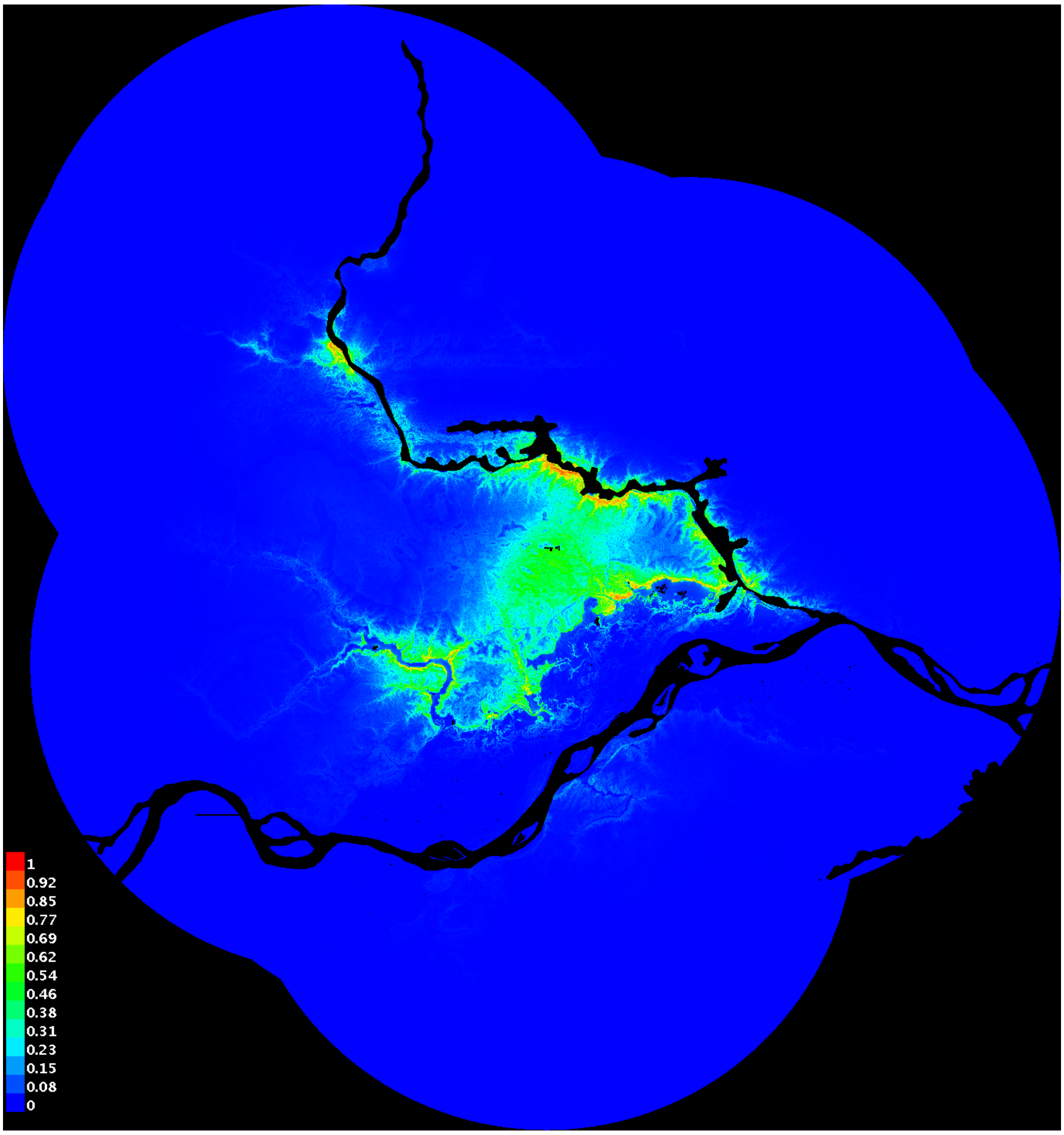


Figure H.11: The mean suitability surface for the Konduri model.

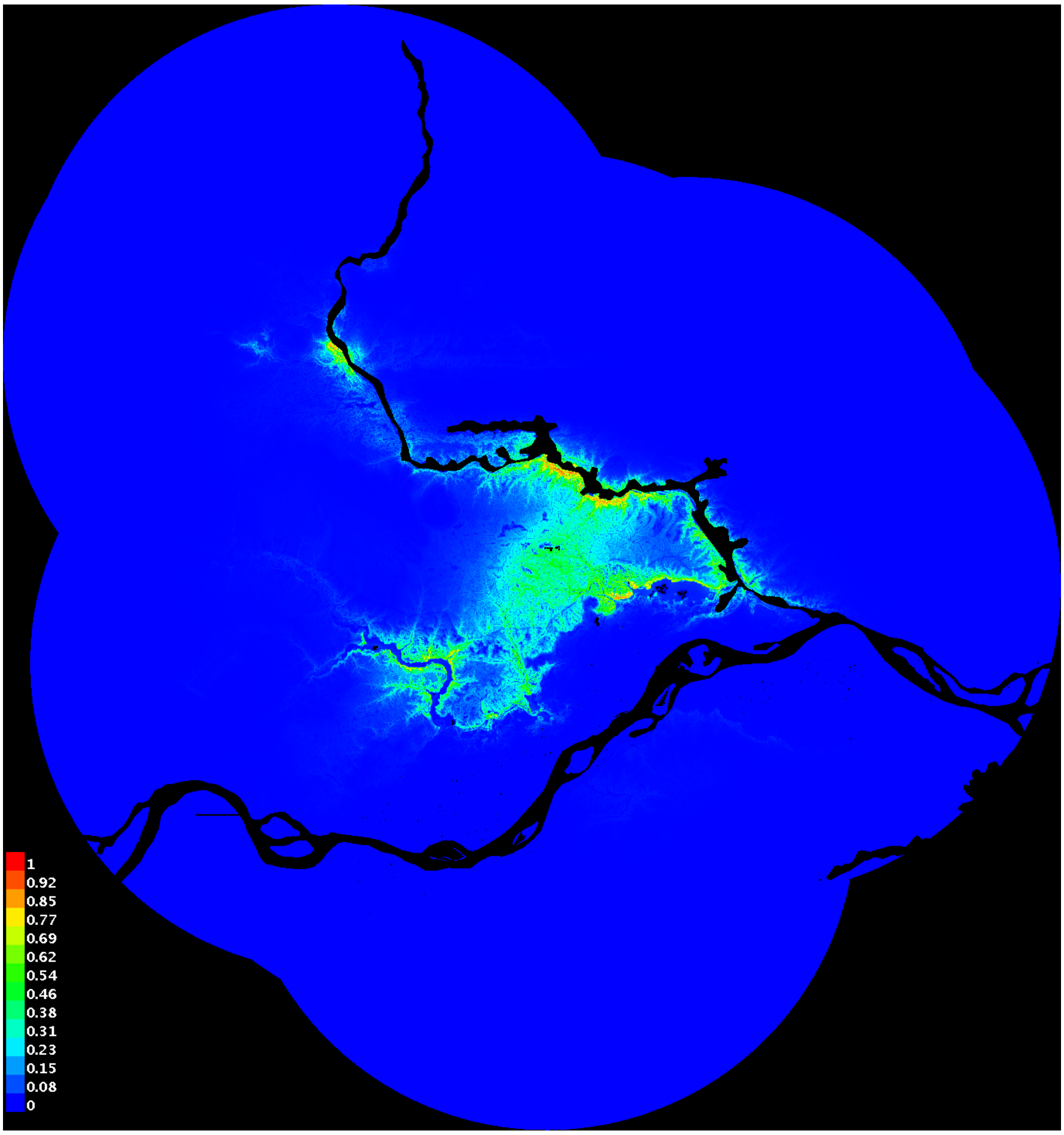


Figure H.12: The minimum suitability surface for the Konduri model.

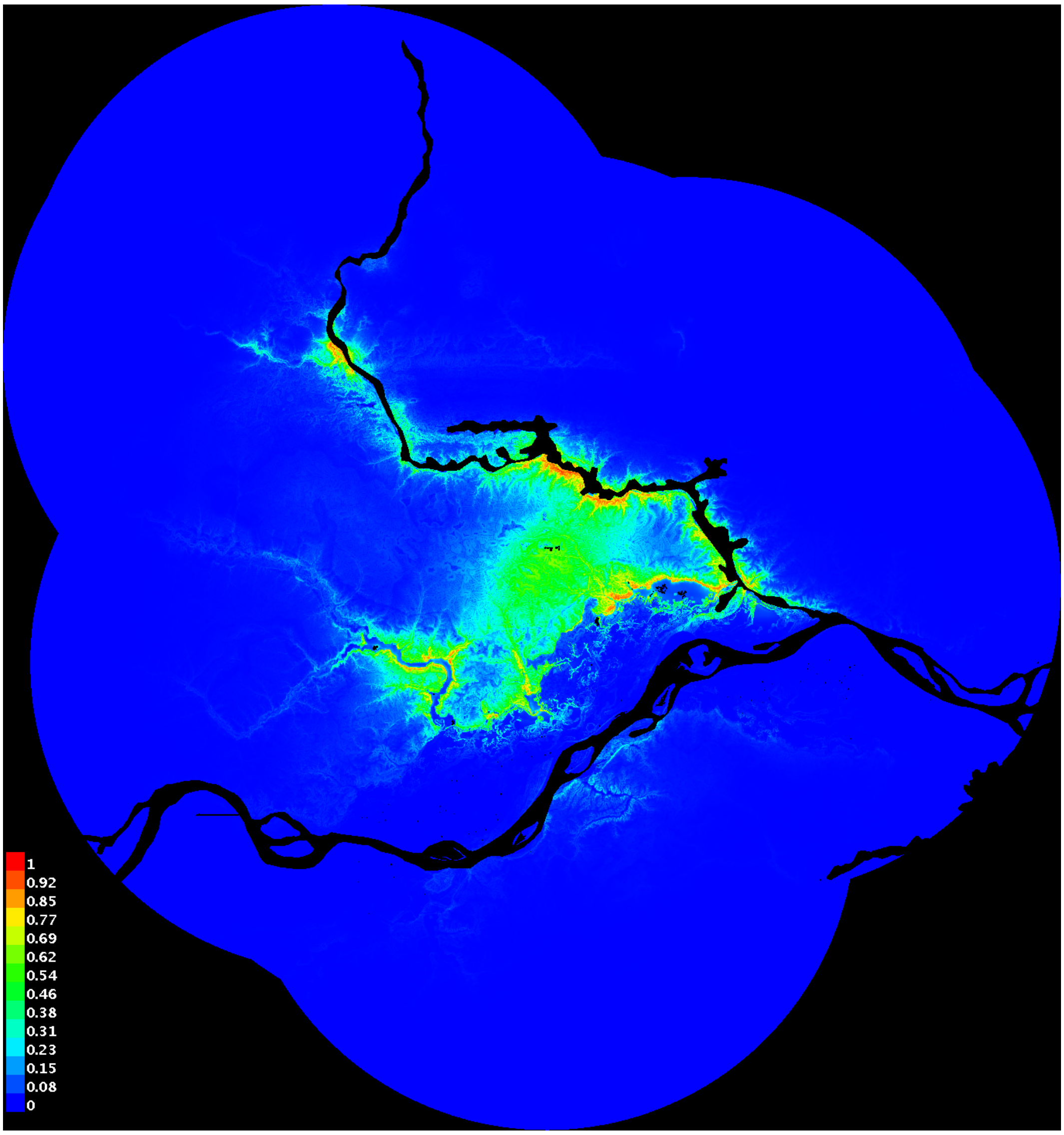


Figure H.13: The maximum suitability surface for the Konduri model.

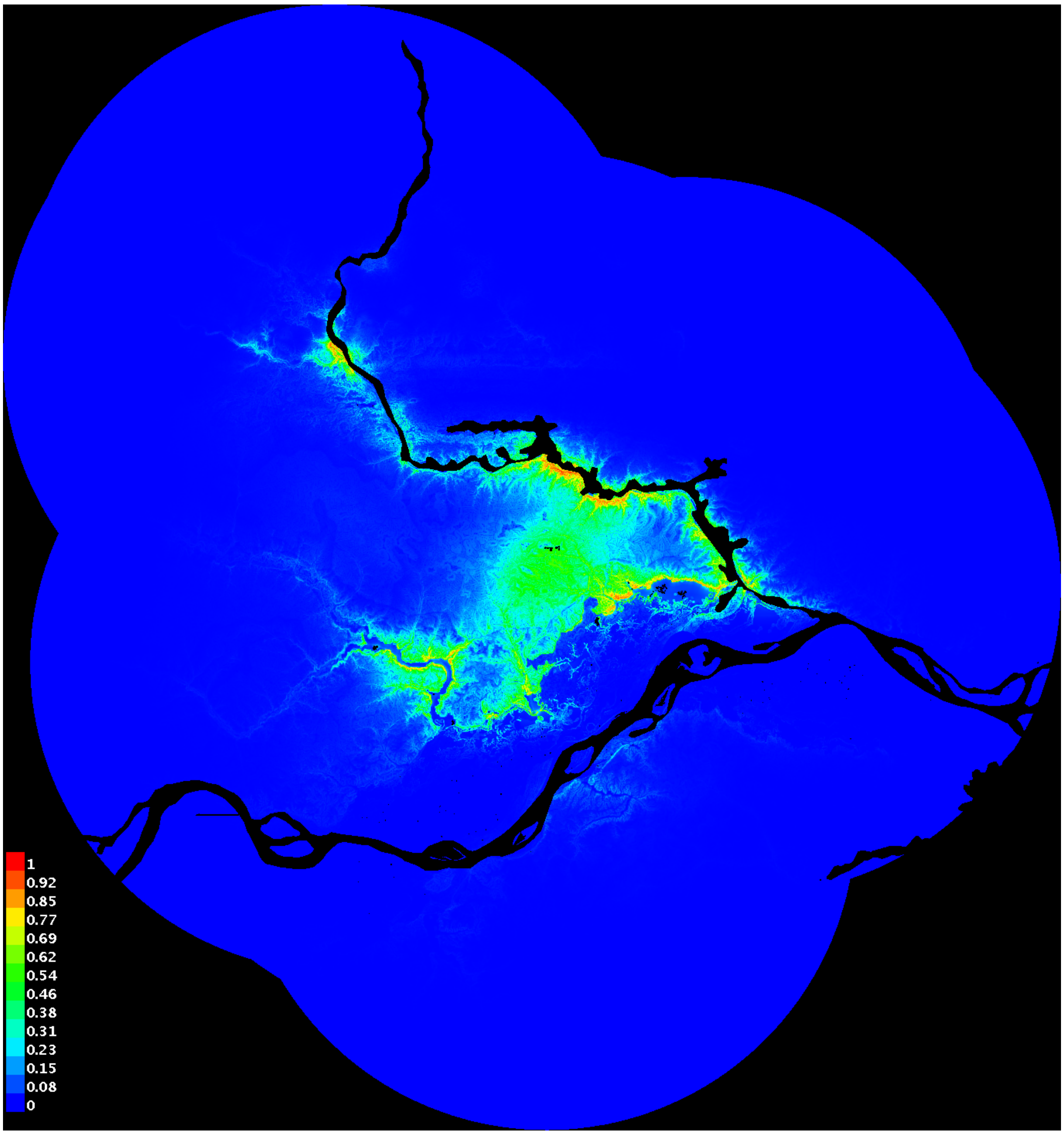


Figure H.14: The median suitability surface for the Konduri model.

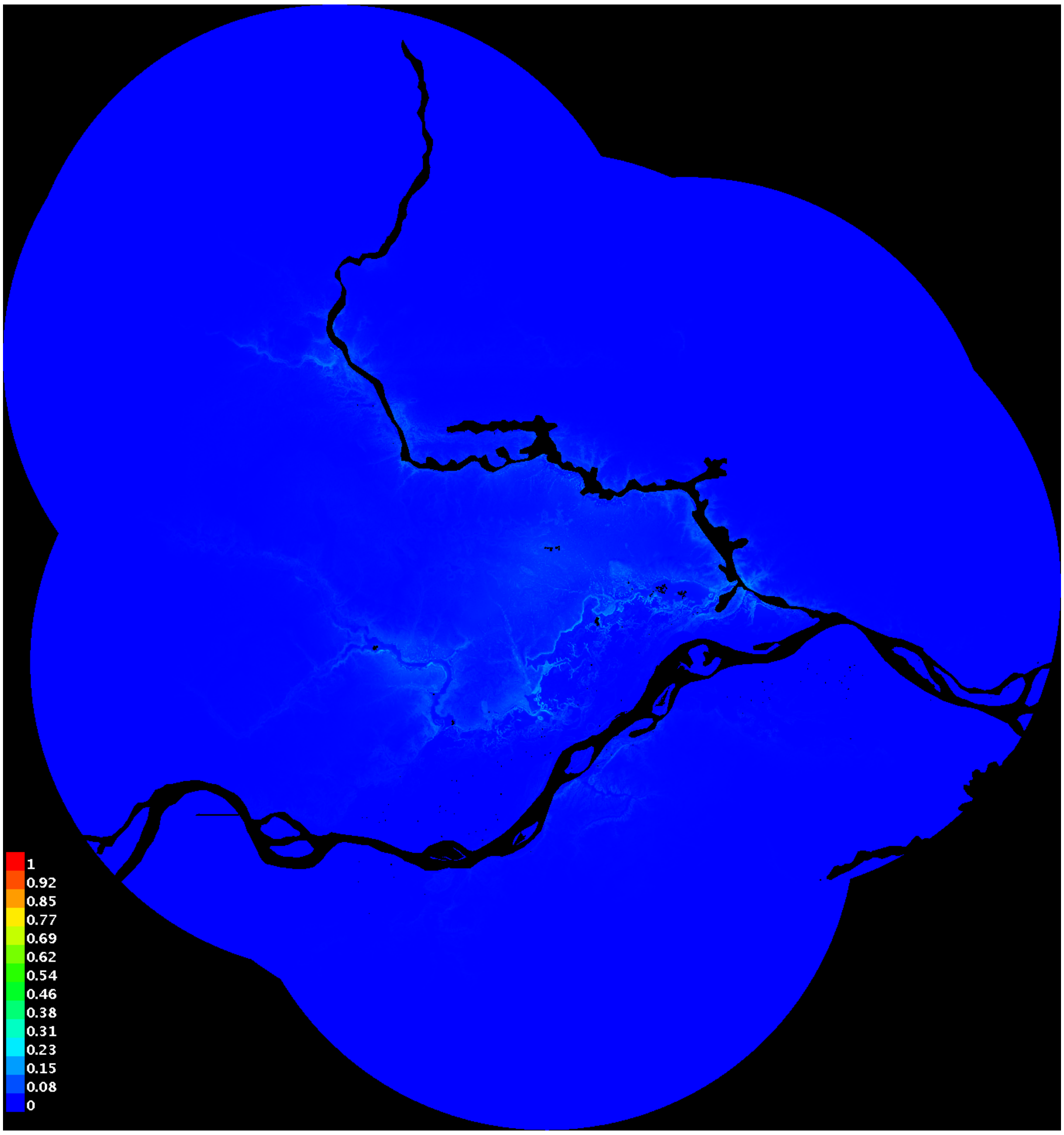


Figure H.15: The standard deviation of the suitability surface for the Konduri model.

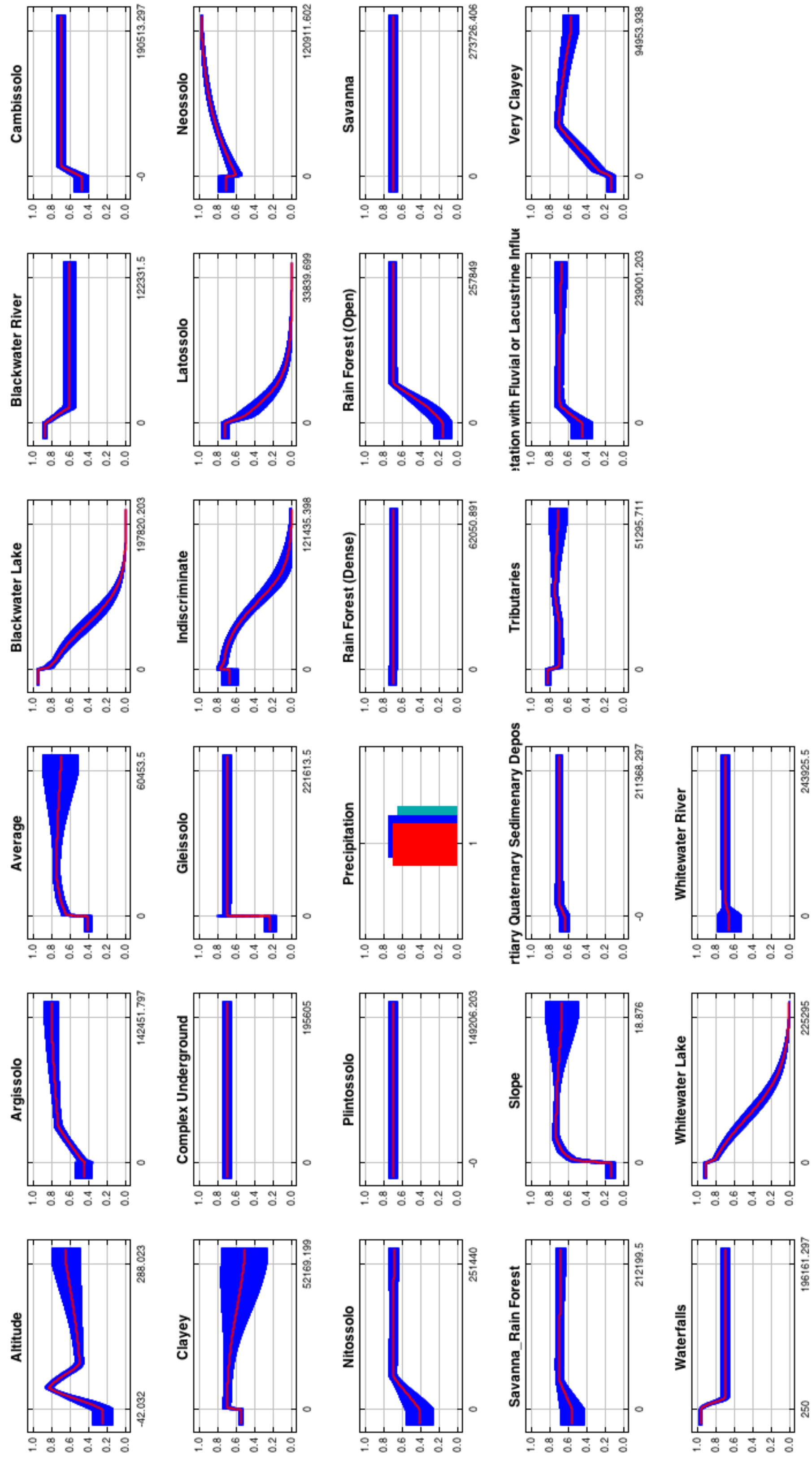


Figure H.16: The marginal response curves for all environmental variables derived using mean values for all but the corresponding variable as model input. This is the result for the Konduri model.

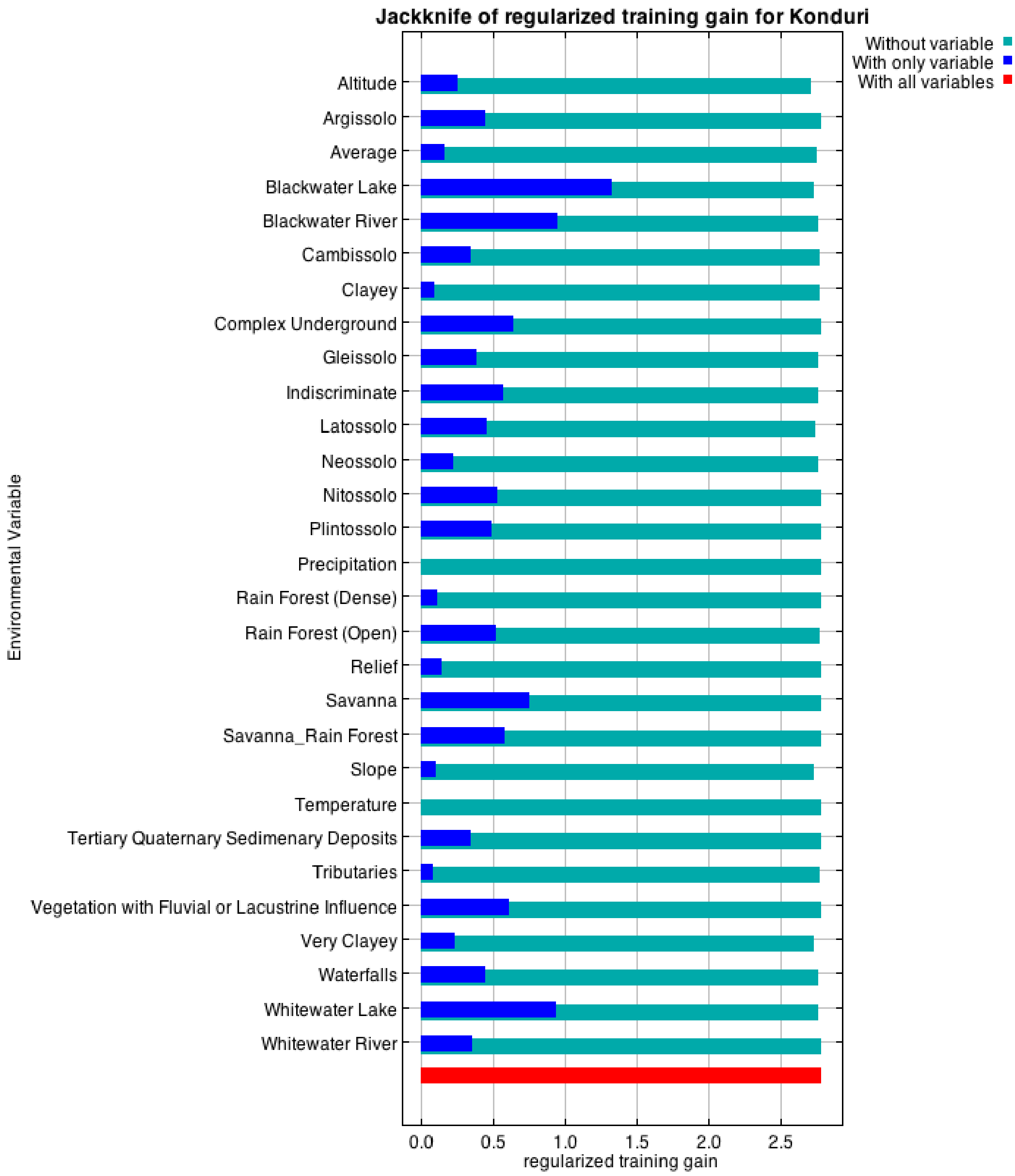


Figure H.17: Jackknife test of regularized training gain for Konduri.

H.2 Guarita

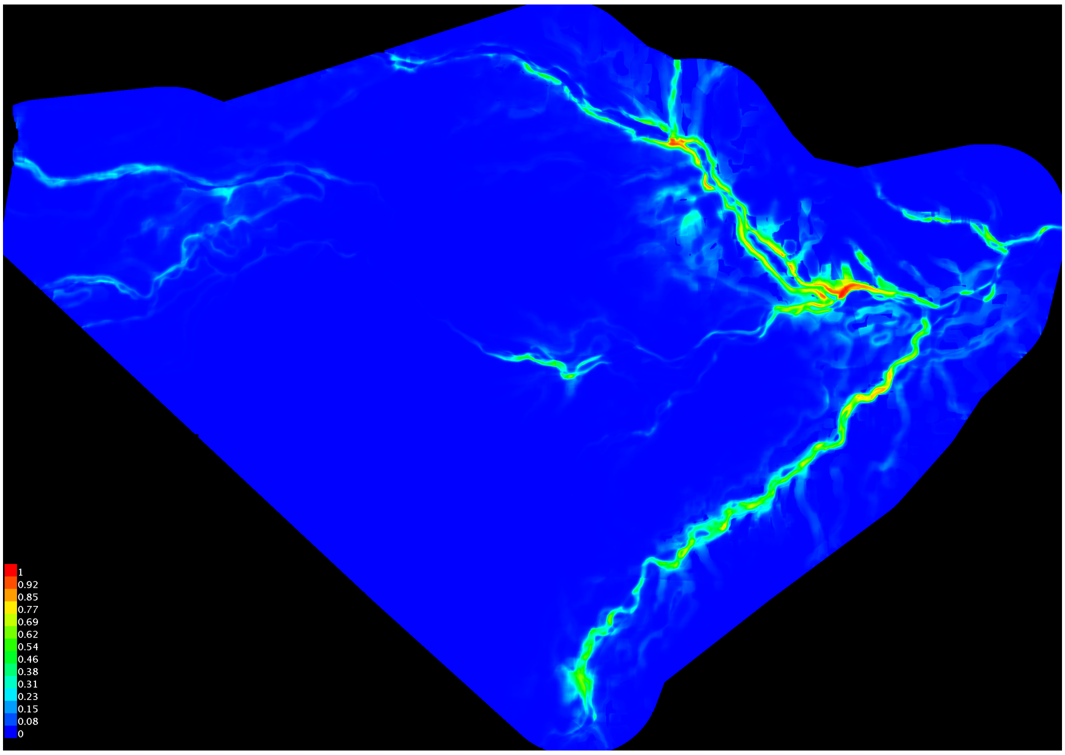


Figure H.28: The mean suitability surface for the Guarita model.

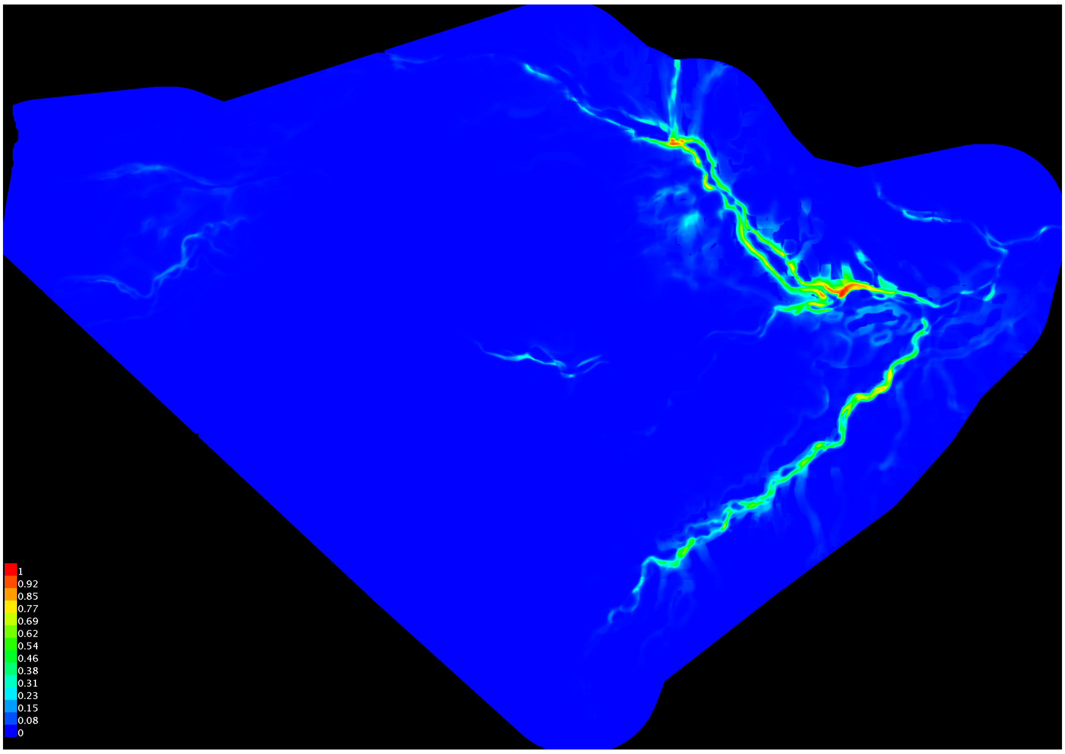


Figure H.29: The minimum suitability surface for the Guarita model.

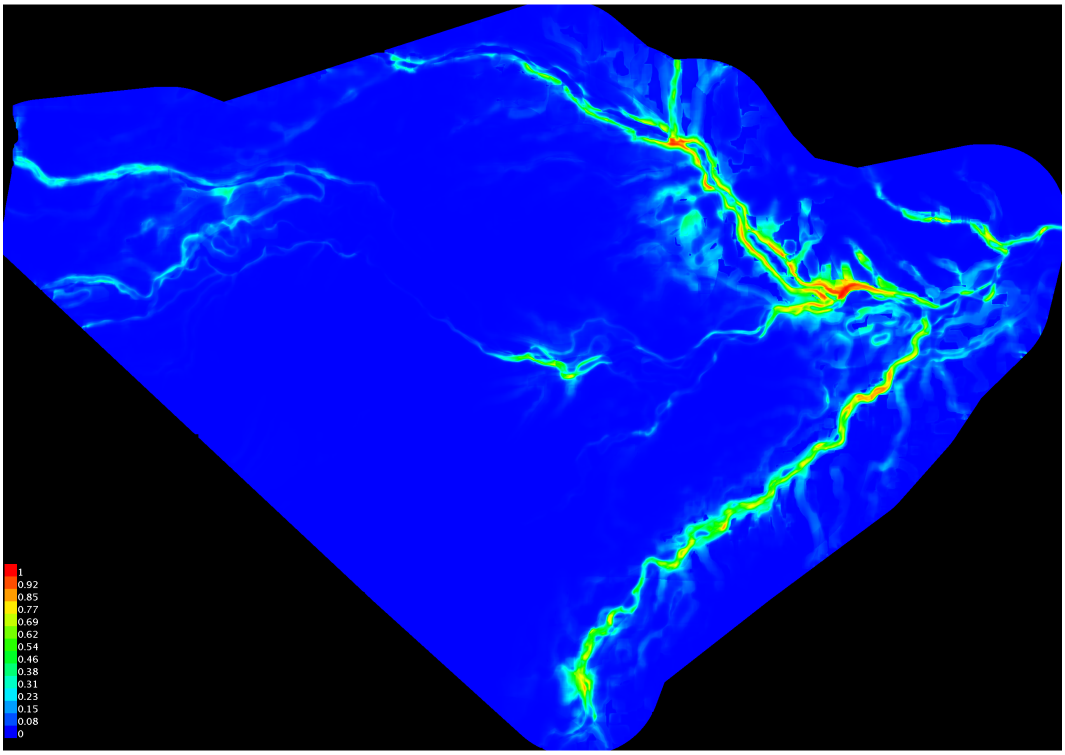


Figure H.210: The maximum suitability surface for the Guarita model.

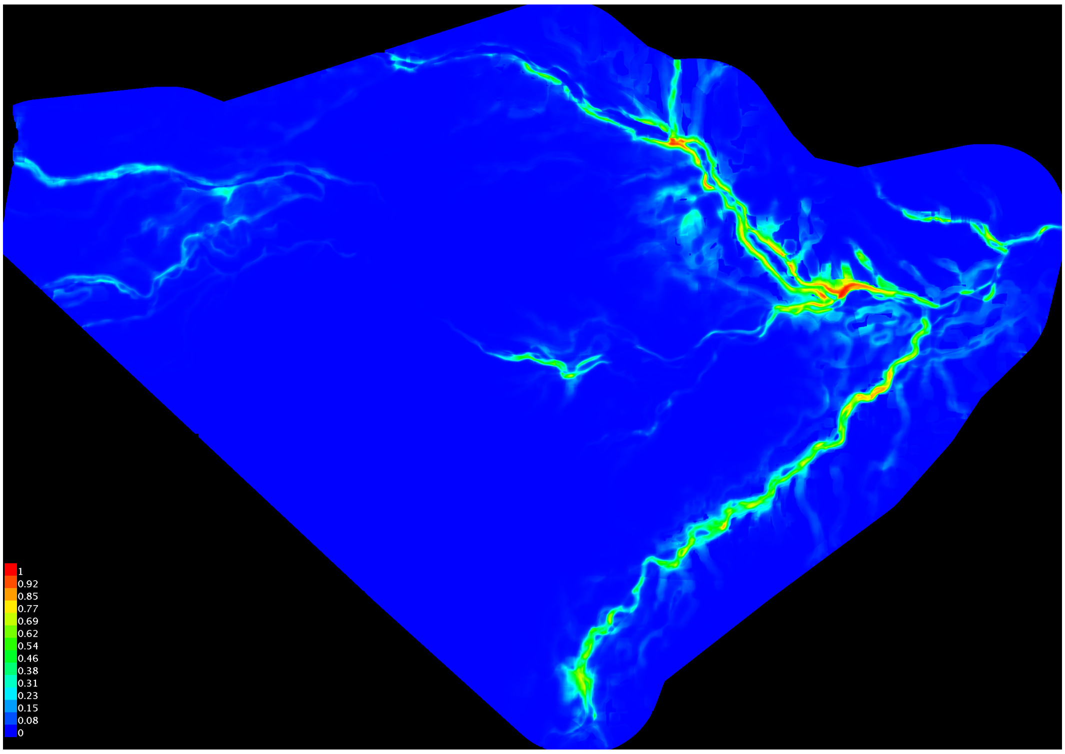


Figure H.211: The median suitability surface for the Guarita model.

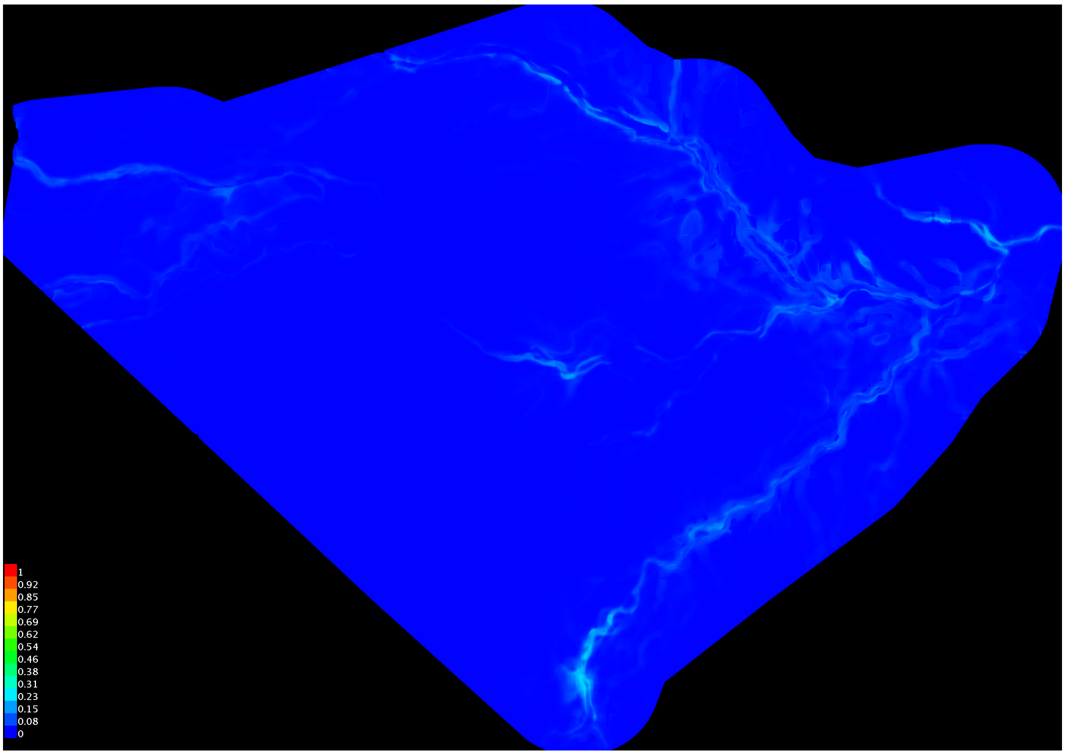


Figure H.212: The standard deviation of the suitability surface for the Guarita model.

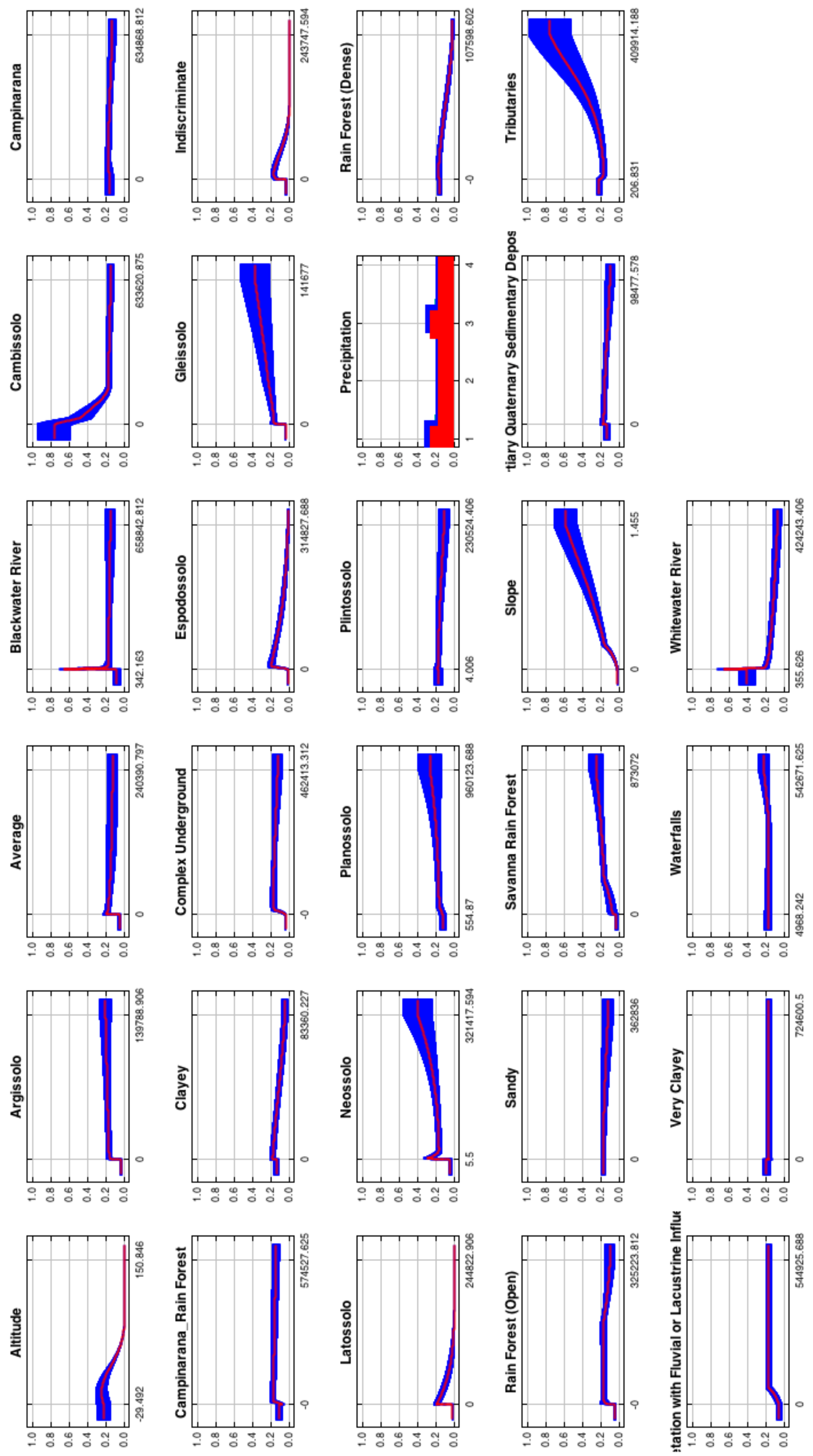


Figure H.213: The marginal response curves for all environmental variables derived using mean values for all but the corresponding variable as model input. This is the result for the Guarita model.

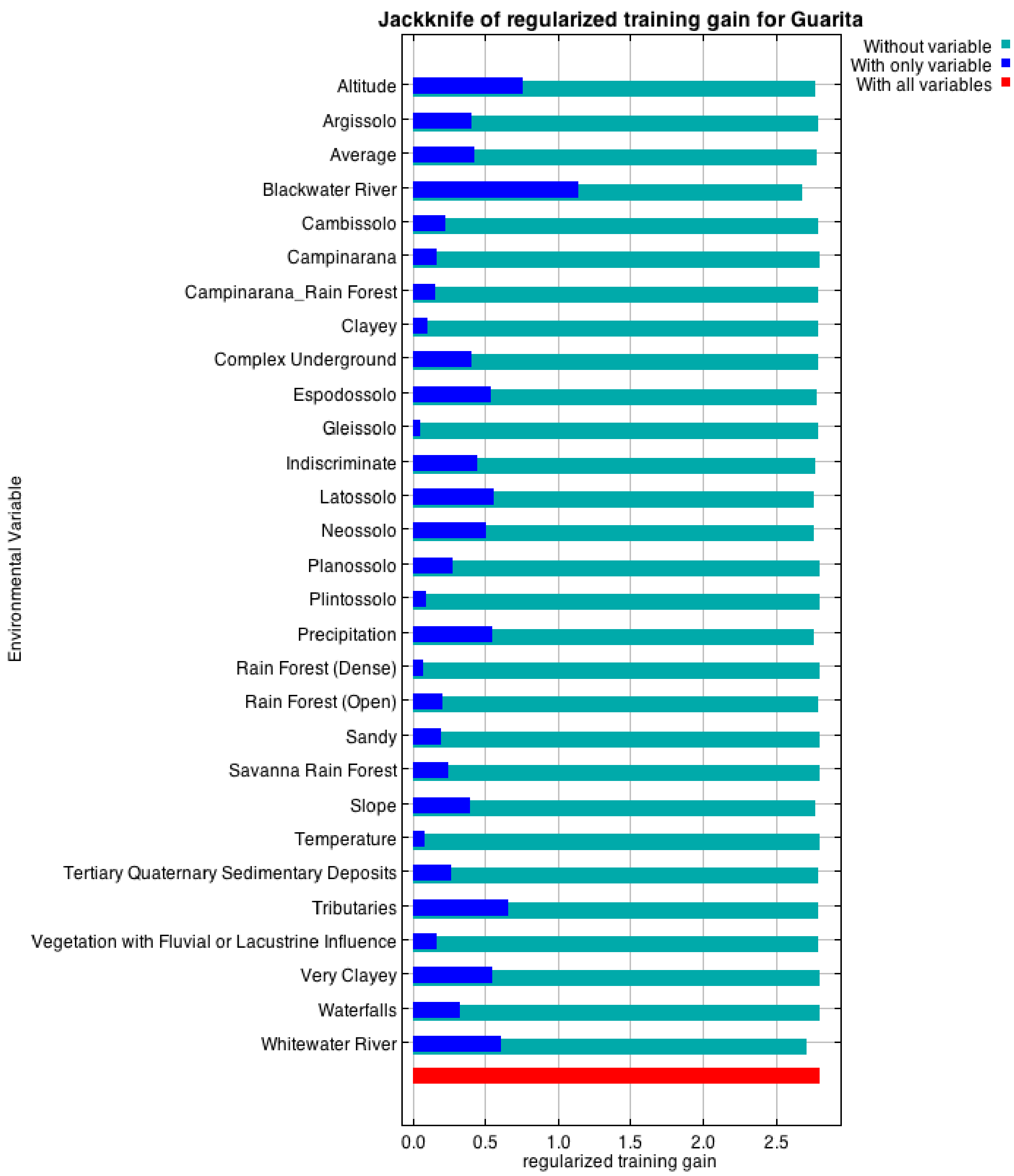


Figure H.214: Jackknife test of regularized training gain for Guarita.

I Environmental Properties

		McMichael	Kvamme	Roper	Judge	Williams
Terrain		Elevation				
		Geological province				
		Roughness (Distance to bluff greater than 25m)				
		Roughness (Distance to bluff greater than 50m)				
		Roughness				
			Slope	Slope	Slope	Slope
Water			Aspect			
		Distance to River		Rivers	Rivers	
Climate	Prec	(Flow accumulation) the size of the neares river from a given pixel				
		Precipitation Seasonality				
		precipitation of the dreist month				
		precipitation of the dreist quarter				
		Precipitation of the wettest month				
	Temp	precipitation of the warmest quarter				
		mean temperature of the dreist quarter				
		Minimum Temperatureof the coldest month				
		Temperature seasonality (Standard Deviation *100)				
Soil		isothermality				
		subsoil bulk density				
		subsoil organic carbon				
		subsoil pH				
		Subsoil cation exchange content				
		subsoil gravel				
View		Topsoil organic carbon				
Protection					View	
Roads			Protection			
Land Use						
Vegetation						
Social						
			Site Density			
			Site Proximity			
Ressources			Spacing			

		Plog	Hurlbett	Parker	Scholtz	Brown	Lovis	Jochim	Grady	Zhang
Terrain										Elevation
		Roughness	Roughness							
										Slope
Water									Aspect	Aspect
				Rivers	Rivers	Rivers	Rivers			Distance to River
Climate	Prec									
	Temp									
Soil										
View										
Protection								View		
Roads								Protection		
Land Use										Distance to Road
Vegetation										
Social										
Ressources										

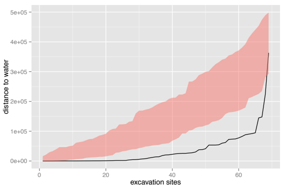
		Goebeö	Gonzalez	Xiao	Wie	Green	Hodder
Terrain		Elevation	Elevation	Elevation		Environment	
						Environment	
						Environment	
						Environment	
					Roughness	Environment	
		Slope	Slope	Slope	Slope	Environment	
						Environment	
Water		Aspect	Aspect	Aspect	Aspect	Environment	
		Distance to River	Distance to River	Distance to River	Hydrology	Environment	Distance to River
Climate	Prec					Environment	
					Climate	Environment	
					Climate	Environment	
					Climate	Environment	
					Climate	Environment	
	Temp				Climate	Environment	
					Climate	Environment	
					Climate	Environment	
					Climate	Environment	
Soil					Climate	Environment	
						Environment	Soil Type
						Environment	Soil Type
						Environment	Soil Type
						Environment	Soil Type
						Environment	Soil Type
View						Environment	Soil Type
Protection							Defense
Roads							
Land Use		Distance to Road	Distance to Road	Distance to Road		Distance to Trade Routes	Proximity to Routes and M
Vegetation					land Use	Environment	
Social					Vegetation	Environment	Vegetation Type
							Location of other sites
							Location of other sites
Ressources							Location of other sites
							Distance to Materials

		Stanic	Egeland
Terrain			Elevation
		Slope	Slope
Water			Aspect
		Distance to Coast	Distance to Rivers
Climate	Prec		
	Temp		
Soil			
View			
Protection		Intervisibility	
Roads			
Land Use		arkets	
Vegetation			
Social			Land Cover
		Location of other sites	
		Location of other sites	
Ressources		Location of other sites	

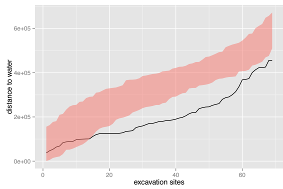
Table I.01: Environmental properties used in archaeological settlement pattern analysis.

J Monte Carlo Simulation

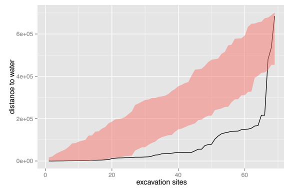
Guarita



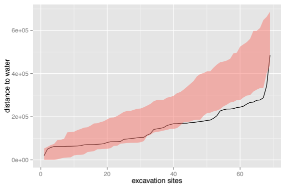
Whitewater River



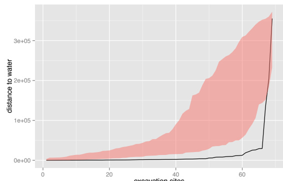
Waterfalls



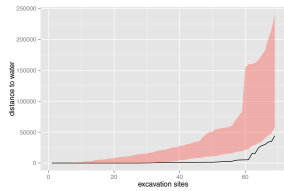
Very Clayey



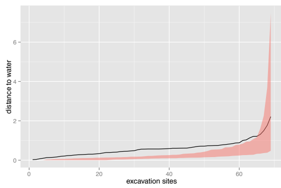
Vegetation with Fluv. or Lac. Inf.



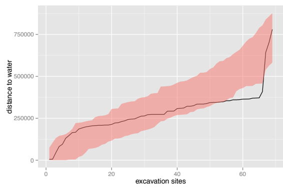
Tributaries



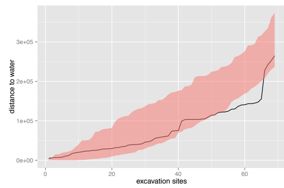
Tert. Quat. Sedimentary Deposits



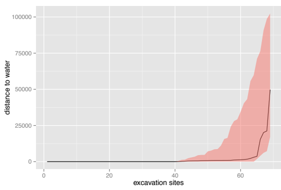
Slope



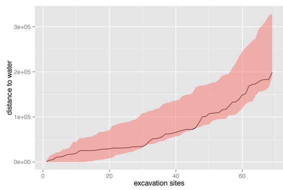
Savanna/Rain Forest



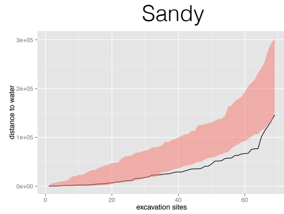
Sandy



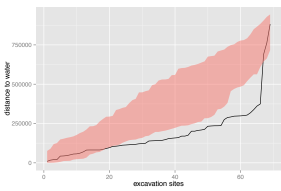
Rain Forest (Open)



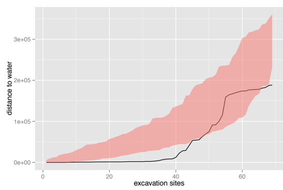
Rain Forest (Dense)



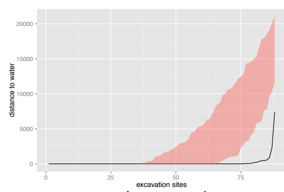
Plintossolo



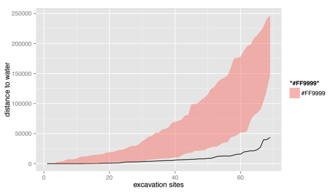
Planossolo



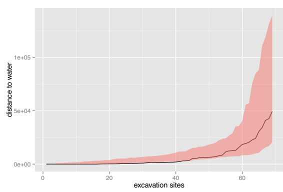
Neossolo



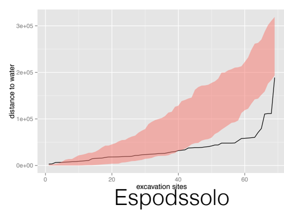
Latossolo



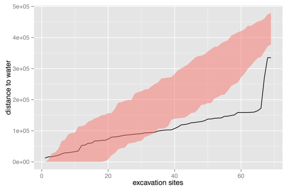
Indiscriminate



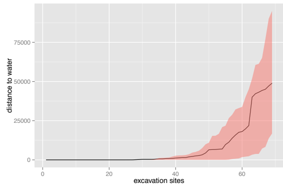
Gleissolo



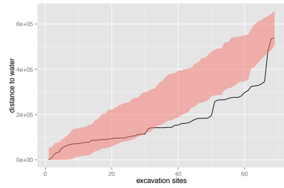
Espodssolo



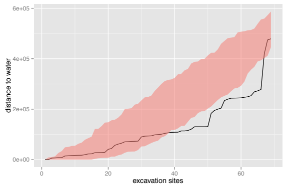
Complex Underground



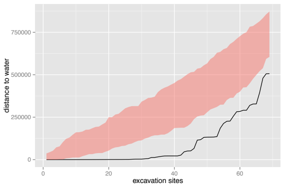
Clayey



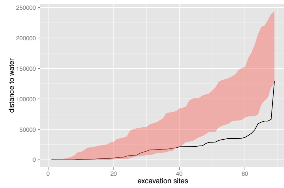
Campinarana/Rain Forest



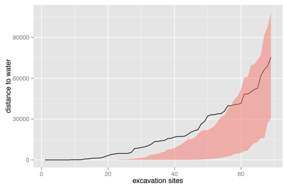
Campinarana



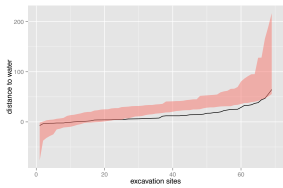
Blackwater River



Average

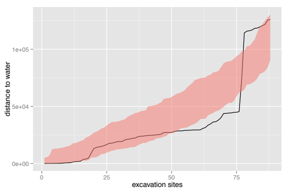


Argissolo

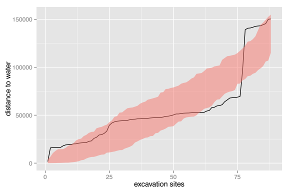


Altitude

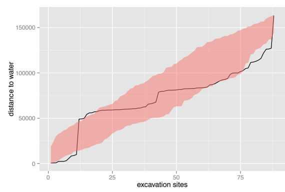
Konduri



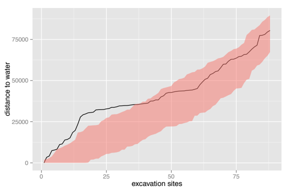
Whitewater River



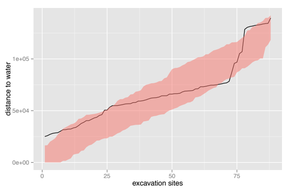
Whitewater Lake



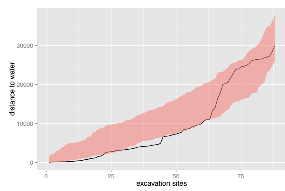
Waterfalls



Very Clayey



Vegetation with Fluv. or Lac. Inf.



Tributaries

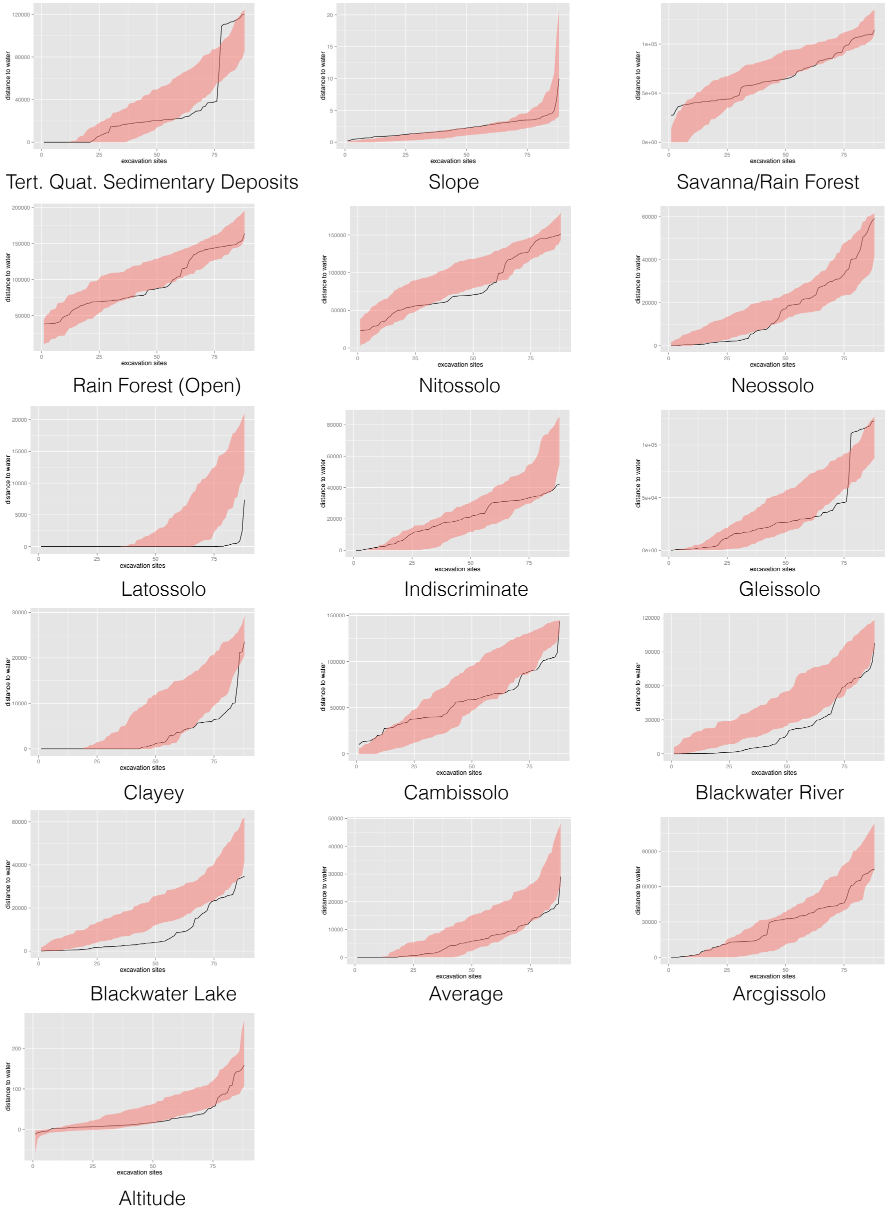
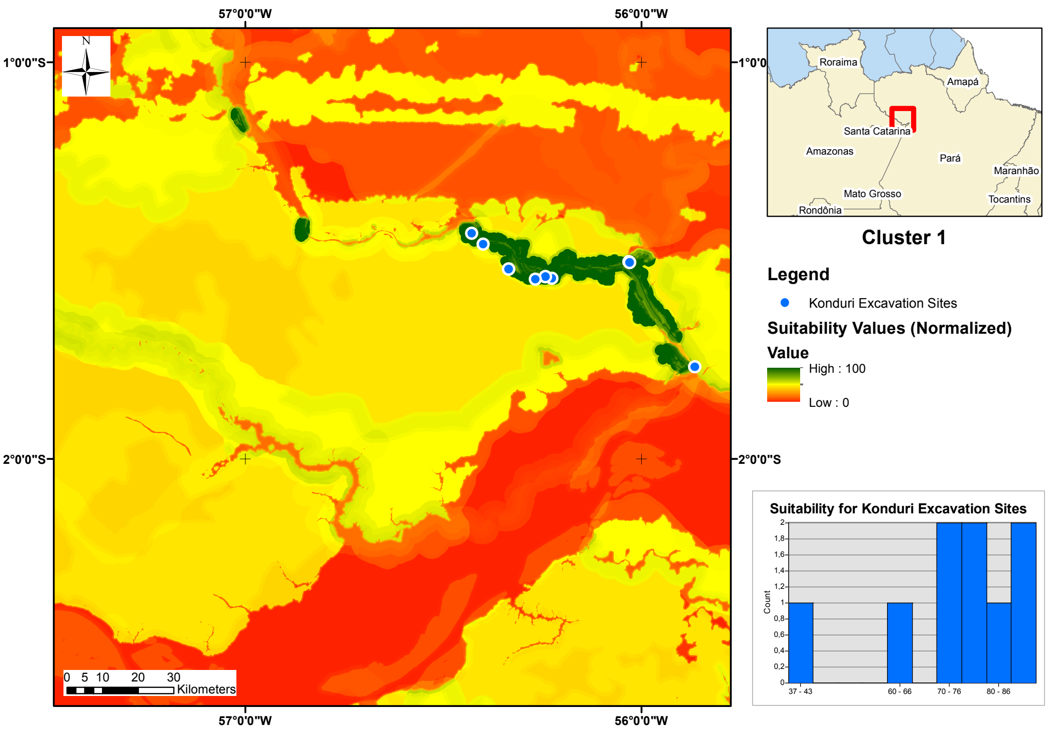
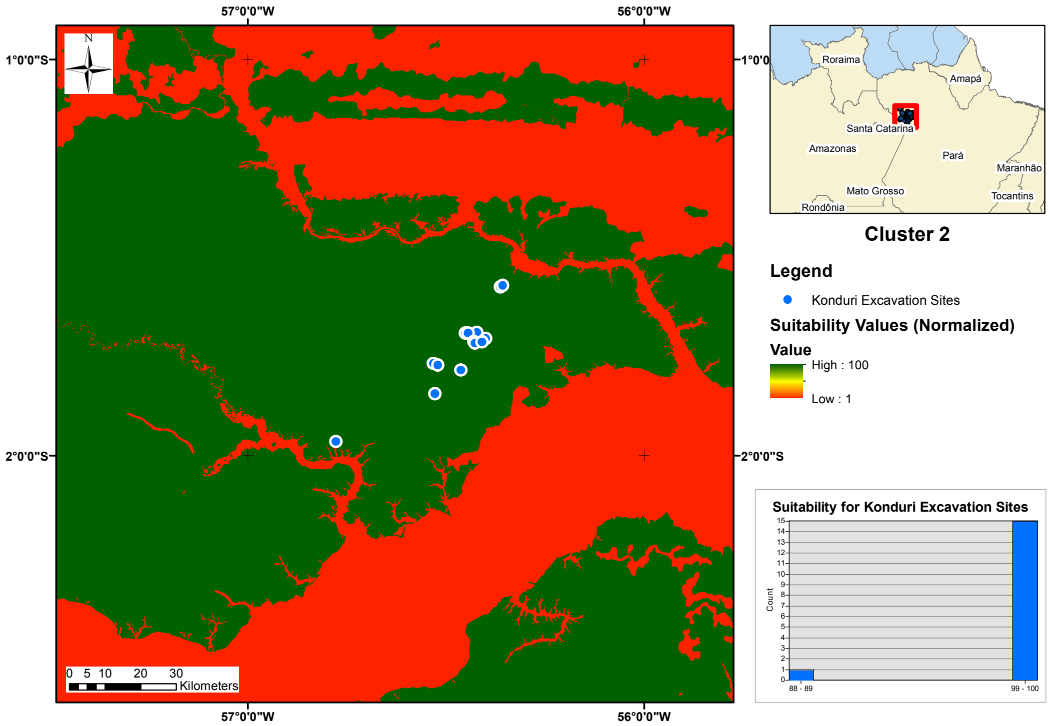


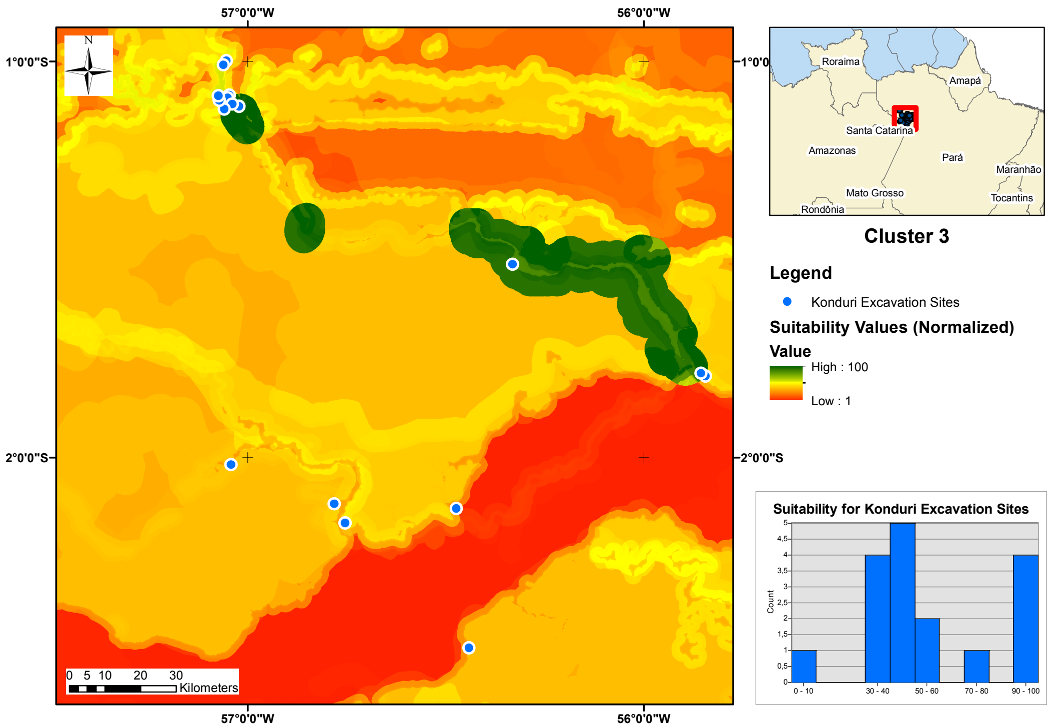
Figure J.01: Results of the Monte Carlo simulation. The excavation sites are shown on the x-axis show and the distance is shown on the y-axis. The black line indicates the distance (sorted in ascending order) from the excavation sites to the environmental parameter whereas the envelope depicts the minimum and maximum distances of the randomly dispersed points.

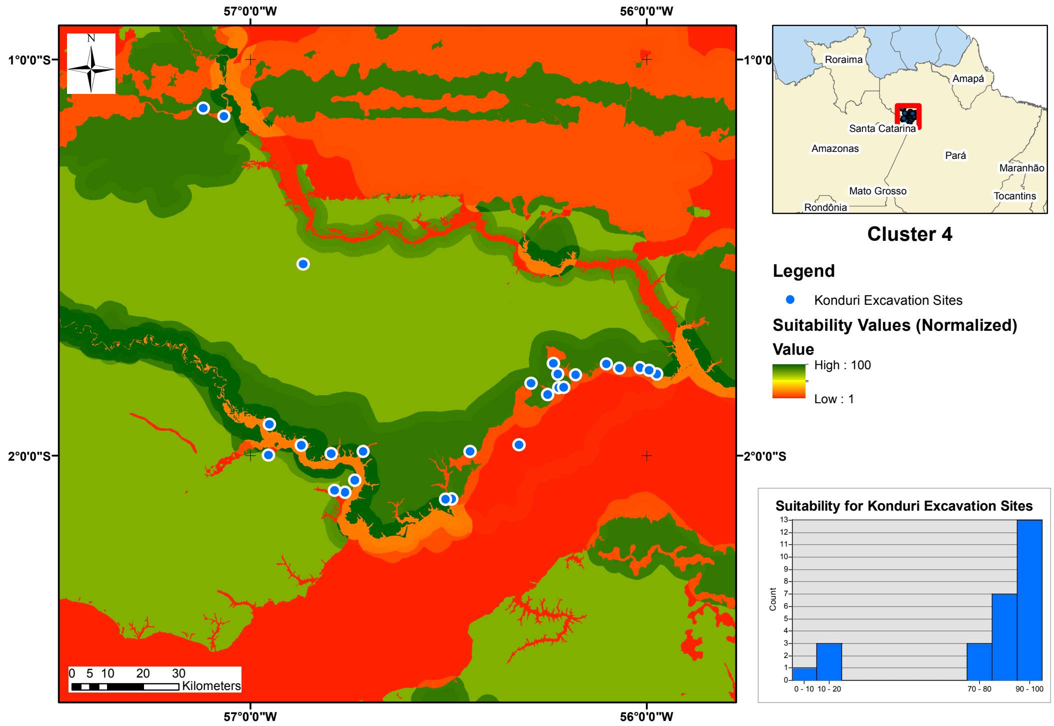
K Suitability Surfaces

K.1 Konduri




57°0'0"W
56°0'0"W
1°0'0"S
2°0'0"S
N
0 5 10 20 30
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 2
Legend
Konduri Excavation Sites
Suitability Values (Normalized)
Value
High : 100
Low : 1
Suitability for Konduri Excavation Sites
Count
88 - 89
99 - 100


57°0'0"W
56°0'0"W
1°0'0"S
2°0'0"S
N
0 5 10 20 30
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 3
Legend
Konduri Excavation Sites
Suitability Values (Normalized)
Value
High : 100
Low : 1
Suitability for Konduri Excavation Sites
Count
0 - 10
30 - 40
50 - 60
70 - 80
90 - 100



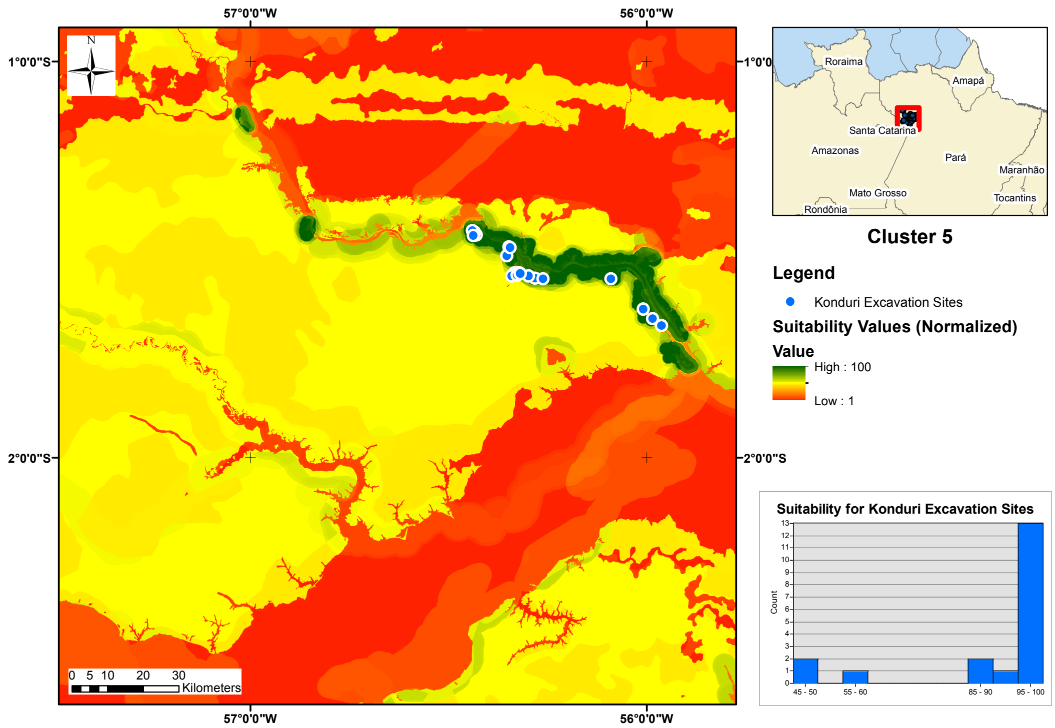
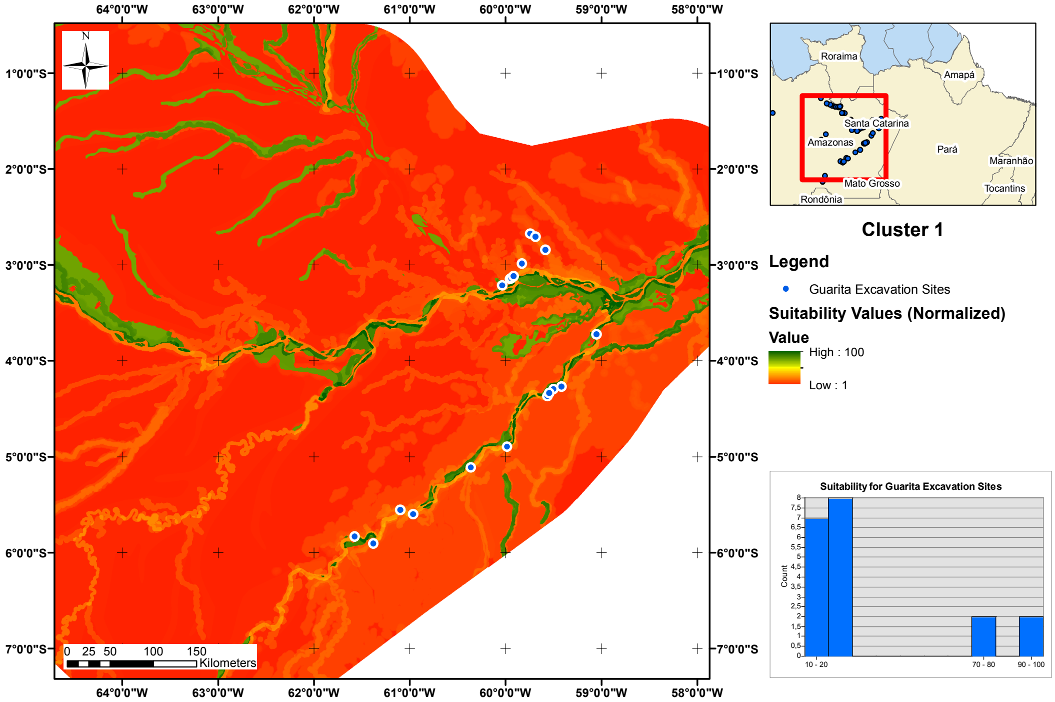
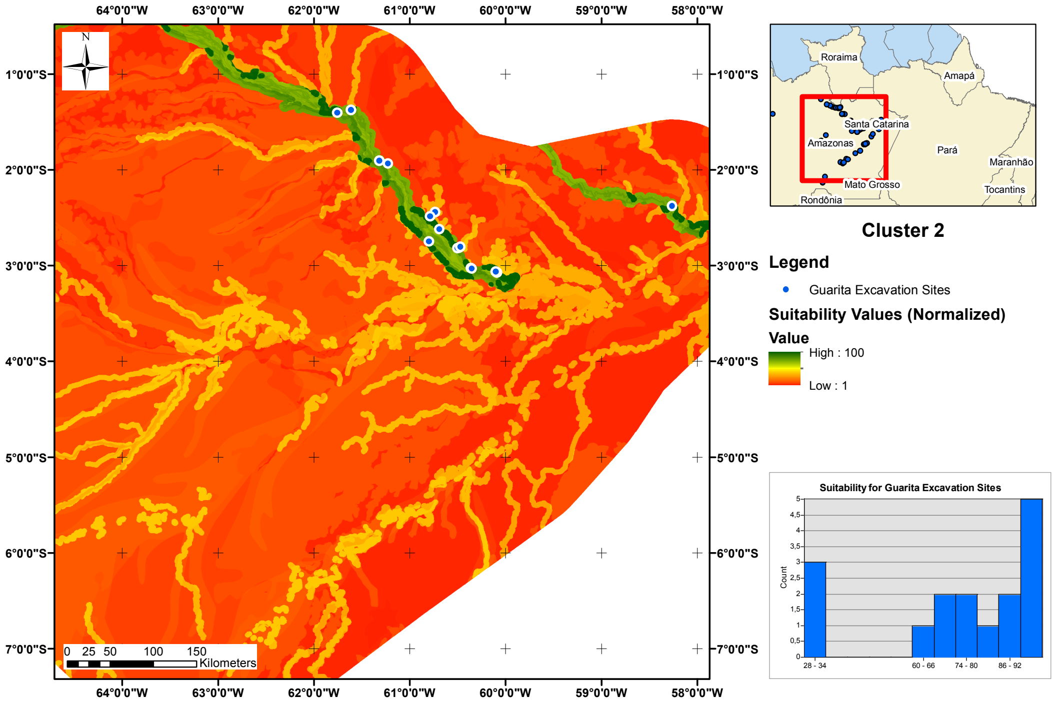


Figure K.11: Suitability surface for cluster 1 – 5 for the Konduri culture. The raster values are normalized whereas a high value indicates a high suitability.

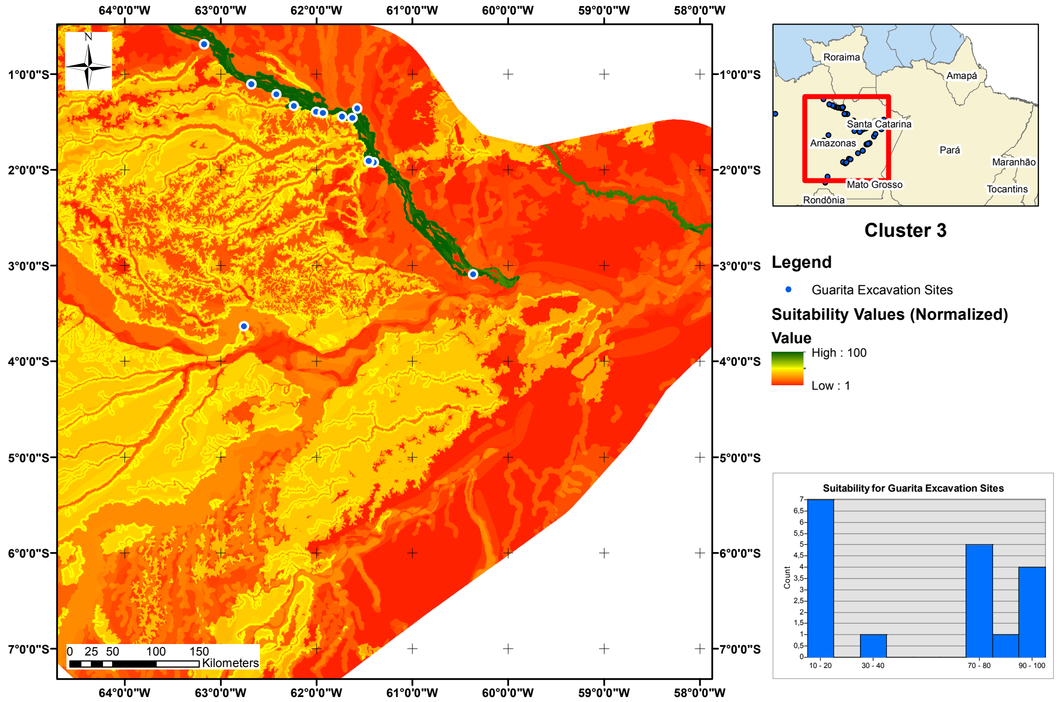
K.2 Guarita

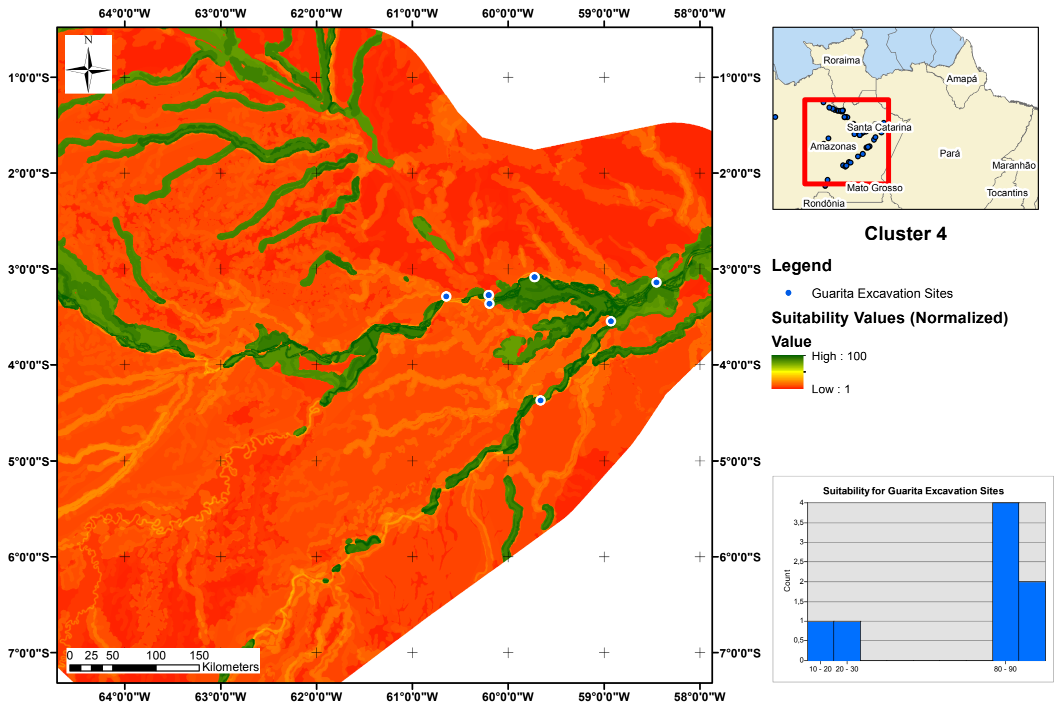


Cluster 1
Legend
Guarita Excavation Sites
Suitability Values (Normalized)
Value
High : 100
Low : 1
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Suitability for Guarita Excavation Sites
Count
10 - 20
70 - 80
90 - 100
0 25 50 100 150
Kilometers
64°0'0"W
63°0'0"W
62°0'0"W
61°0'0"W
60°0'0"W
59°0'0"W
58°0'0"W
1°0'0"S
2°0'0"S
3°0'0"S
4°0'0"S
5°0'0"S
6°0'0"S
7°0'0"S



Cluster 2
Legend
Guarita Excavation Sites
Suitability Values (Normalized)
Value
High : 100
Low : 1
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Suitability for Guarita Excavation Sites
Count
28 - 34
60 - 66
74 - 80
86 - 92
0 25 50 100 150
Kilometers
64°0'0"W
63°0'0"W
62°0'0"W
61°0'0"W
60°0'0"W
59°0'0"W
58°0'0"W
1°0'0"S
2°0'0"S
3°0'0"S
4°0'0"S
5°0'0"S
6°0'0"S
7°0'0"S


64°0'0"W
63°0'0"W
62°0'0"W
61°0'0"W
60°0'0"W
59°0'0"W
58°0'0"W
1°0'0"S
2°0'0"S
3°0'0"S
4°0'0"S
5°0'0"S
6°0'0"S
7°0'0"S
N
0 25 50 100 150
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 3
Legend
Guarita Excavation Sites
Suitability Values (Normalized)
Value
High : 100
Low : 1
Suitability for Guarita Excavation Sites
Count
10 - 20
30 - 40
70 - 80
90 - 100


64°0'0"W
63°0'0"W
62°0'0"W
61°0'0"W
60°0'0"W
59°0'0"W
58°0'0"W
1°0'0"S
2°0'0"S
3°0'0"S
4°0'0"S
5°0'0"S
6°0'0"S
7°0'0"S
N
0 25 50 100 150
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 4
Legend
Guarita Excavation Sites
Suitability Values (Normalized)
Value
High : 100
Low : 1
Suitability for Guarita Excavation Sites
Count
10 - 20
20 - 30
80 - 90

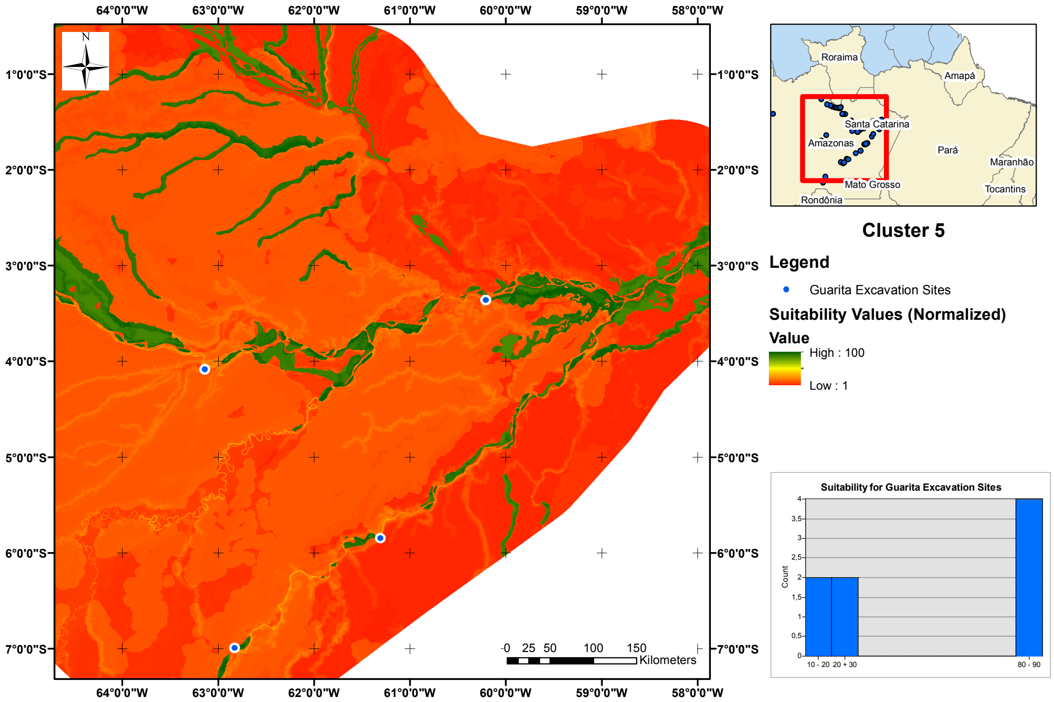
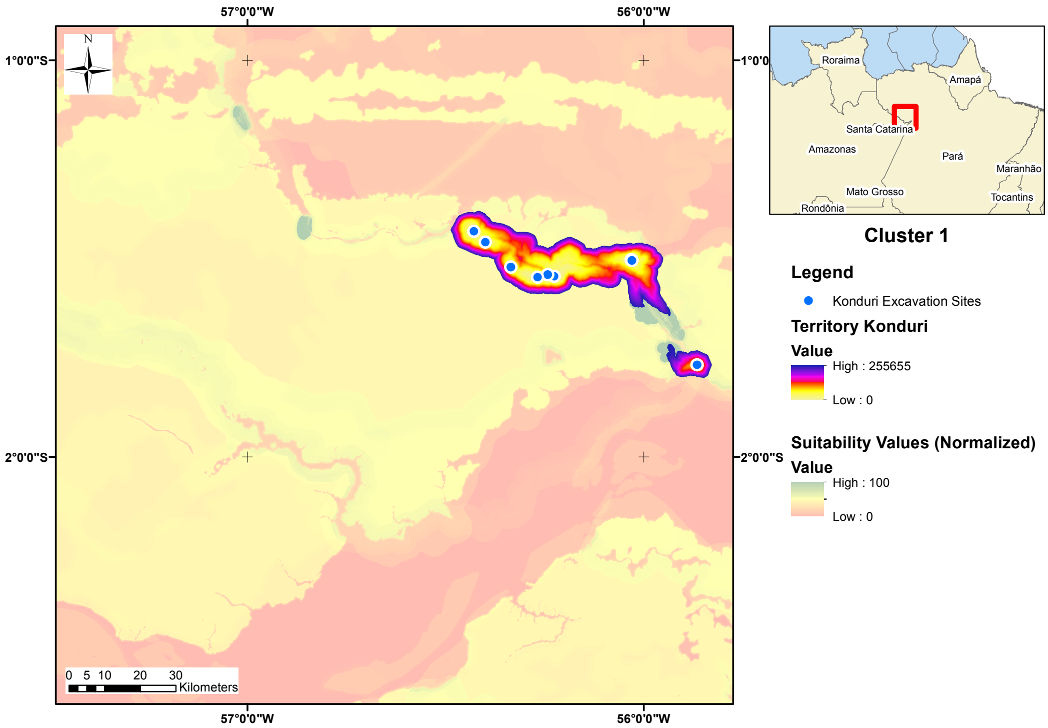
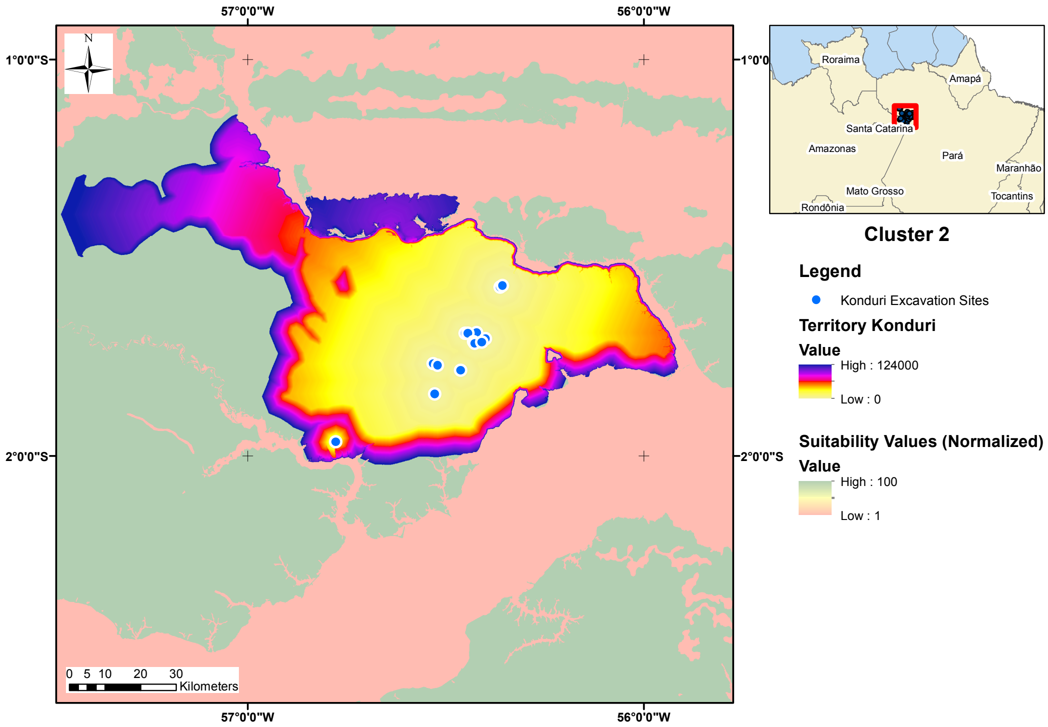


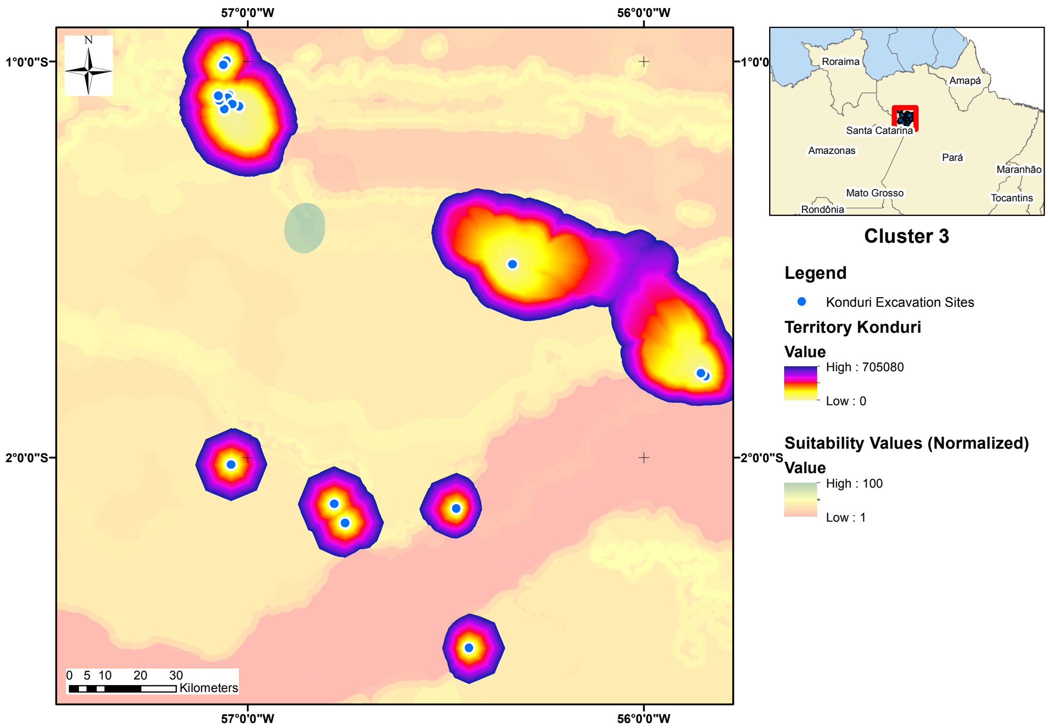
Figure K.22: Suitability surface for cluster 1 – 5 for the Guarita culture. The raster values are normalized whereas a high value indicates a high suitability.

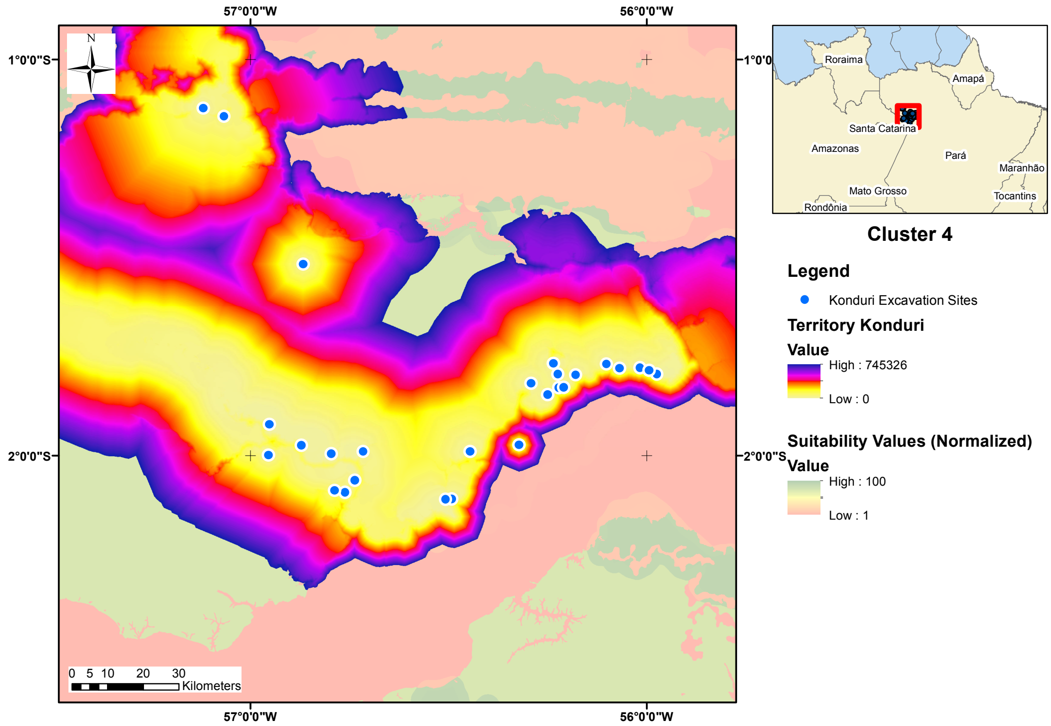
L Territory Analysis

L.1 Konduri




57°0'0"W
56°0'0"W
1°0'0"S
2°0'0"S
N
0 5 10 20 30
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 2
Legend
Konduri Excavation Sites
Territory Konduri
Value
High : 124000
Low : 0
Suitability Values (Normalized)
Value
High : 100
Low : 1


57°0'0"W
56°0'0"W
1°0'0"S
2°0'0"S
N
0 5 10 20 30
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 3
Legend
Konduri Excavation Sites
Territory Konduri
Value
High : 705080
Low : 0
Suitability Values (Normalized)
Value
High : 100
Low : 1



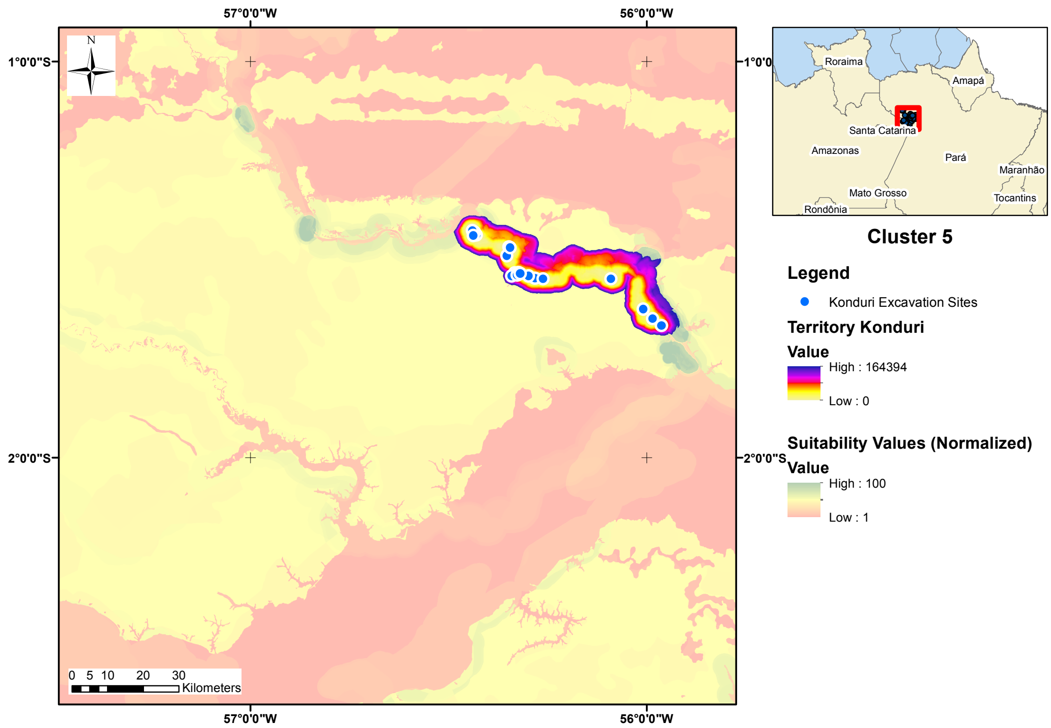
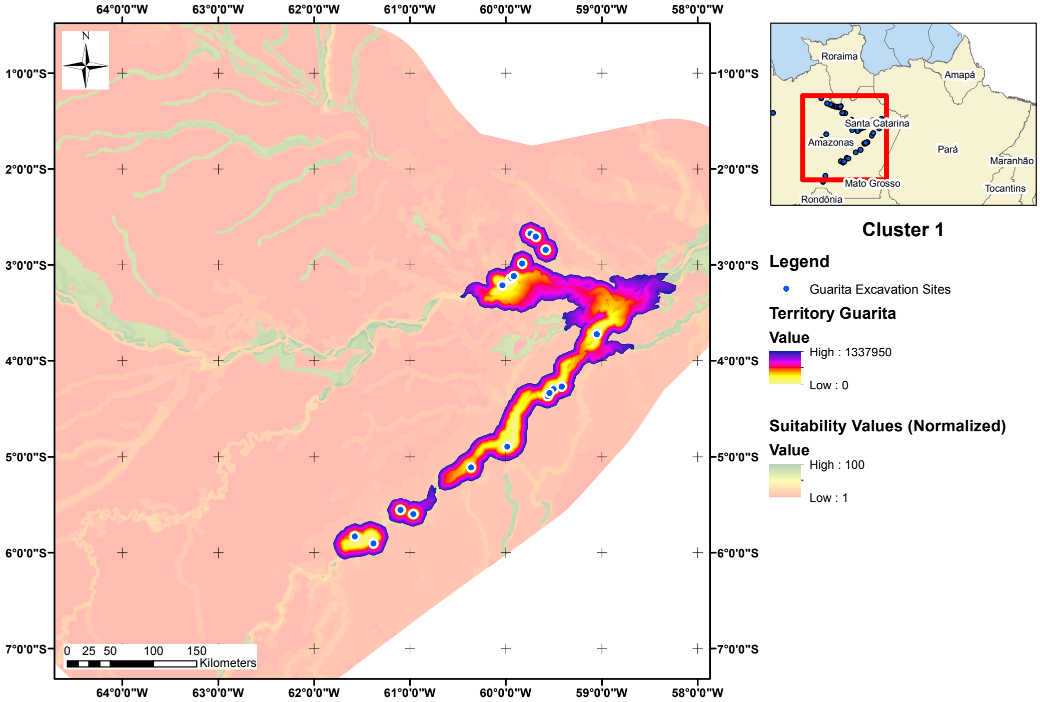
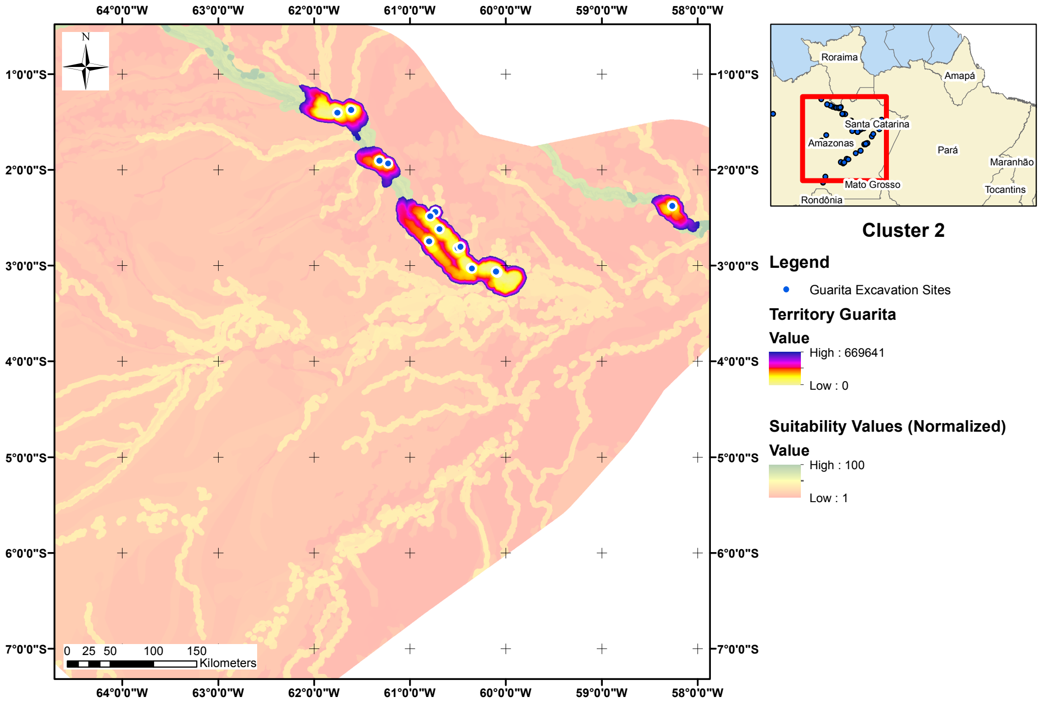
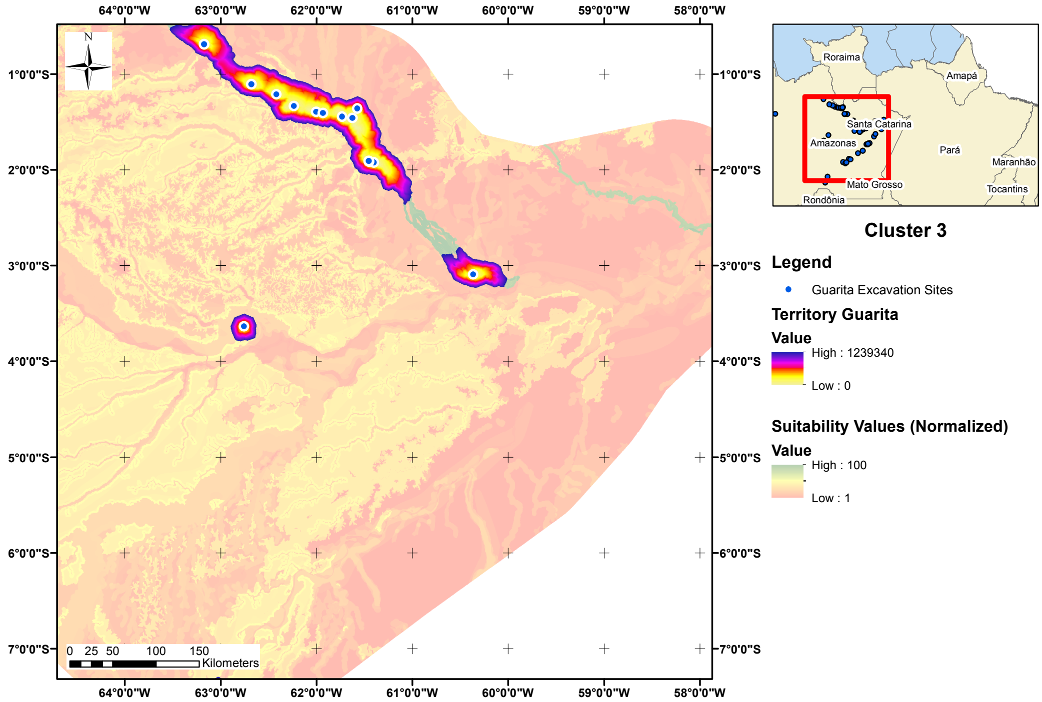


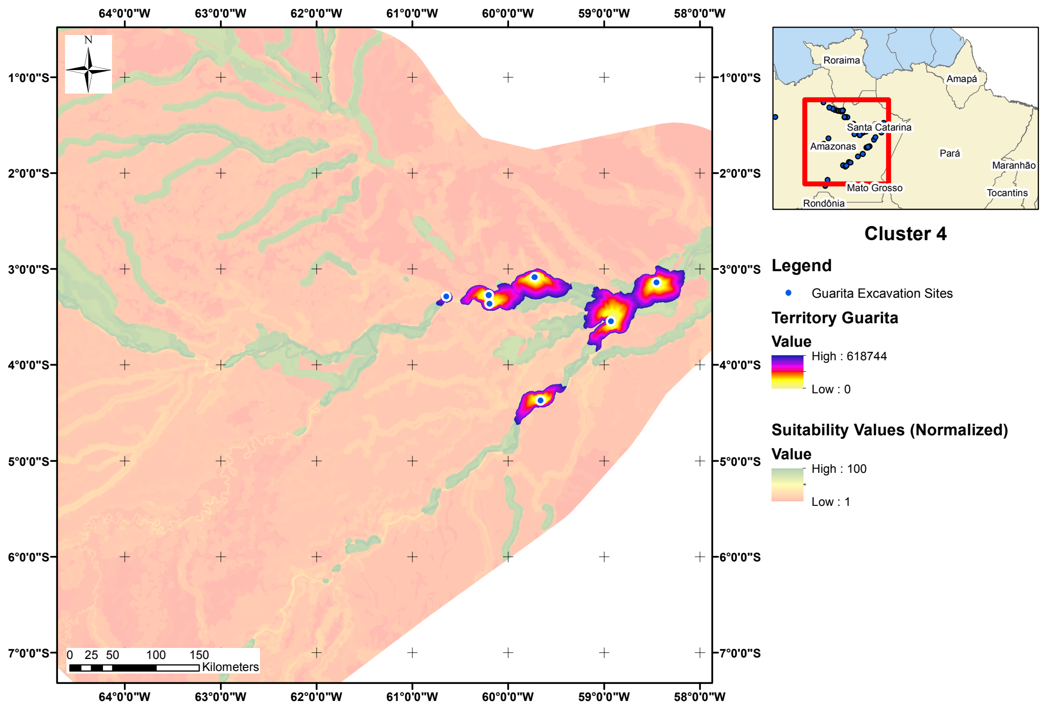
Figure L.11: Territory for cluster 1 – 5 for the Konduri culture. The raster values are based on CD_{max} whereas a high value indicates a higher cost.

L.2 Guarita


64°0'0"W
63°0'0"W
62°0'0"W
61°0'0"W
60°0'0"W
59°0'0"W
58°0'0"W
1°0'0"S
2°0'0"S
3°0'0"S
4°0'0"S
5°0'0"S
6°0'0"S
7°0'0"S
N
0 25 50 100 150
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 1
Legend
Guarita Excavation Sites
Territory Guarita
Value
High : 1337950
Low : 0
Suitability Values (Normalized)
Value
High : 100
Low : 1


64°0'0"W
63°0'0"W
62°0'0"W
61°0'0"W
60°0'0"W
59°0'0"W
58°0'0"W
1°0'0"S
2°0'0"S
3°0'0"S
4°0'0"S
5°0'0"S
6°0'0"S
7°0'0"S
N
0 25 50 100 150
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 2
Legend
Guarita Excavation Sites
Territory Guarita
Value
High : 669641
Low : 0
Suitability Values (Normalized)
Value
High : 100
Low : 1


64°0'0"W
63°0'0"W
62°0'0"W
61°0'0"W
60°0'0"W
59°0'0"W
58°0'0"W
1°0'0"S
2°0'0"S
3°0'0"S
4°0'0"S
5°0'0"S
6°0'0"S
7°0'0"S
N
0 25 50 100 150
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 3
Legend
Guarita Excavation Sites
Territory Guarita
Value
High : 1239340
Low : 0
Suitability Values (Normalized)
Value
High : 100
Low : 1


64°0'0"W
63°0'0"W
62°0'0"W
61°0'0"W
60°0'0"W
59°0'0"W
58°0'0"W
1°0'0"S
2°0'0"S
3°0'0"S
4°0'0"S
5°0'0"S
6°0'0"S
7°0'0"S
N
0 25 50 100 150
Kilometers
Roraima
Amapá
Santa Catarina
Amazonas
Pará
Maranhão
Mato Grosso
Tocantins
Rondônia
Cluster 4
Legend
Guarita Excavation Sites
Territory Guarita
Value
High : 618744
Low : 0
Suitability Values (Normalized)
Value
High : 100
Low : 1

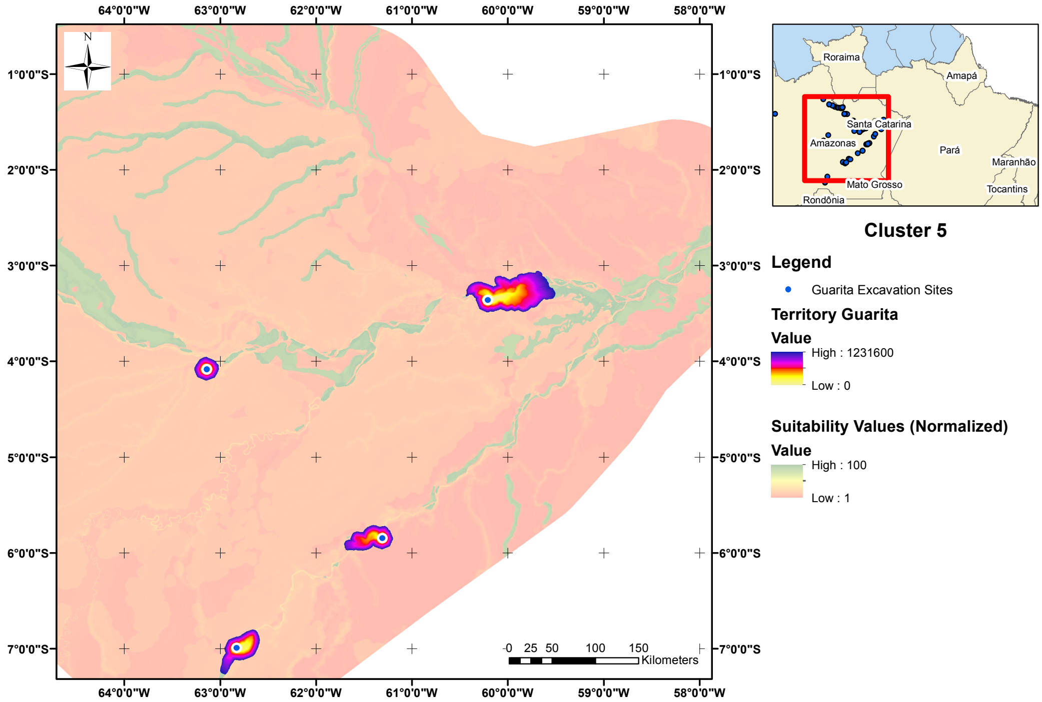


Figure L.22: Territory for cluster 1 – 5 for the Guarita culture. The raster values are based on CD_{max} whereas a high value indicates a higher cost.