Interactive WebGIS for Archaeological Settlement Pattern Analysis – A Requirement Analysis

Carolin von Groote-Bidlingmaier*, Klaus Hilbert, Johannes Schwer*, Sabine Timpf***

* University of Augsburg, Geoinformatics Group, Germany
** University of Porto Alegre (PUCRS), FFCH, Brazil

Abstract. Only very little is known about prehistoric settlement structures and movement behavior in Amazonia. On the one hand archaeological data is inherently vague which leads to loss of information. On the other hand the lack of a consistent data base to collect, store and/or provide the collected data of former settlements complicates the research. In this research we analyze the known data sources to create the biggest known collection of excavation sites in the Amazon Basin. A data base schema was created to store and provide the collected information. Besides that a requirement analysis based on expert knowledge and literature review was done, to figure out the needs for visual analytics approaches. This application should provide statistical and analytical functions to perform visual analytics without any special knowledge about GIS. The goal of this research is to provide a collection of data and necessary tools to analyze pre-historical data in the Amazon.

Keywords: WebGIS, Visual Analytics, Explorative Analysis, Terra Preta

1. Introduction

In contrast to the high cultures in the Andes (e.g. Inka or Maya) there is only very little information about the cultures in Amazonia. Humans live in the Amazon region for more than 10500 years (Roosevelt et al. 1996). According to the known excavation sites and radiocarbon data the first cultures that were not hunter and gatherers but farmers and herders settled about 6000 years ago.

Hilbert (1968) divided the findings – based on the different pottery styles – in four different traditions which can be subdivided into several cultures.

This classification is used in this paper to distinguish between different population groups. Due to the long time period and the size of the observed region it can be assumed, that certain settlement patterns and movement behavior vary according to the tradition or culture.

In order to examine the settlement behavior of pre-historical cultures in Amazonia a lot of potential influences need to be taken into account which are described later. One reason to choose a visual analytics approach is the complexity of such an analysis. The other reason is that no further GIS or geoinformatic knowledge is required.

Problem is that there is no complete database available. Iphan (Instituto do Patrimônio Histórico e Artístico Nacional) provides a huge collection of excavation sites because it is the official database. It contains the information about the findings such as the culture or tradition that settled there. The disadvantage is that they don't provide any coordinates.

Another collection of excavation sites is provided by Winkler Prins & Aldrich (2010) who tried to locate Amazonian dark earth (ADE) sites using the information of 60 publications and created a file (a google kml.file) which can be used for further analysis. The ADE (also known as terra preta) is a fertile anthropogenic soil which indicates former settlements. Besides the location some other information like the source and the accuracy is given. Unfortunately no information about the official excavation site number or name (which can then be used to match the Iphan database with the location of this dataset) is given, which makes it difficult – if not impossible – to merge this dataset with other data sources. It's been even more difficult because the accuracy of the location varies

To gain knowledge about pre-colonial cultures in the Amazon, further research is required. One aspect is to factor geographical information because it is not an isotropic and homogenous space and therefore the varying environmental circumstances had influence on the decision making processes. O'Sullivan & Unwin (2010) stated that there is no consistent definition of the term spatial analysis. Generally the four topics spatial data manipulation, spatial data analysis, spatial statistical analysis and spatial modeling are associated with spatial analysis, but using the term in different ways. That is why they introduce the term geographic information analysis which is defined as "investigating the patterns that arise as a result of processes that may be operating in space.". A similar definition is given by Mitchell (1999) who says that data analysis in the context of geographical information systems means to look for relationships and geographical patterns between geoobjects. Both definitions imply some activities such as choosing the analysis method, preparing the data or interpret the obtained results. Especially the selection of appropriate methods, the preparation of the data

and the application of the chosen methods require expert knowledge concerning geographical information systems (GIS). Archaeologists do not necessarily have GIS skills and web-based analysis applications might simplify the consideration of geographical influences and therefore can be useful for a broad group of researchers.

This paper is structures as followed. In chapter 2 the general approach of visual analytics and its usage related to archaeology is explained. Chapter 3 lists the required data and tools to analyze settlement patterns in Amazonia. The implemented (but yet incomplete) web GIS application is described in chapter 4. Chapter 5 presents the conclusion.

2. Visual Analytics and Archaeologists

The analysis of spatial influences on the decision making process (i.e. settlement behavior) varies depending on the scale, the availability of data, or the scope. Due to the inherently vague data an exploratory approach might be useful to generate hypothesis in the beginning. The term exploratory data analysis (EDA) was introduced by Tukey (1977). He invented this philosophy of how analyzing data should be performed. The idea of EDA is, to detect and describe patterns, trends, and relationships, thus generate hypothesis (Andrienko & Andrienko 2006). Although Tukey didn't precisely defined the term he described its characteristics (Badie et al. 2011). This description includes three main strategies of data analysis, namely graphical presentation, provision of flexibility in viewpoint and in facilities, and intensive search for parsimony and simplicity (Jones 1986). Although this concept initially was not invented for spatial data but emerged from statistics, the idea can be matched to geographical approaches. That is due to the fact that EDA is "strongly associated with the use of graphical representations of data." (Andrienko & Andrienko 2006). But not only EDA is useful when analyzing archaeological data but also confirmatory analysis. Both two approaches plus a third, namely presentation, are the main goals of visualizing data (Keim et al. 2008). In terms of representing spatial data visualization and cartography have a lot in common (O'Sullivan & Unwin 2010) and maps are helpful way to visualize spatial distributions or patterns.

The term visual analytics describes an iterative process that combines computational tools with the human "capabilities to perceive, relate and conclude" (Keim et al. 2008). Thus, visual analytics includes visualization but also involves other research topics such as knowledge discovery or statistical analysis (see Figure 1). In terms of geographical information the core of visual analytics focuses on space and time (Andrienko et al. 2010). Spatio-temporal analyses don't necessarily have to be performed by experts but is something which is done by everybody (i.e. when planning a journey etc.). The focus has to be on developing and implementing tools which leads to interactive or dynamic visual interfaces to facilitate the analytical reasoning (Thomas 2005).

What does that mean regarding to archaeological or more precisely pre-colonial cultural studies in the Amazon? The research about settlements of pre-colonial cultures is related to the distribution of ADEs (Smith 1980). Although the origin of this anthrosol is not yet clarified, it is known that the cultures – intentional or unintentional – produced the dark earths to increase the carrying capacity. That means that these terra preta sites always indicate a former settlement. Thus a broad knowledge about the production workflow and therefore the environmental requirements is necessary. Besides the soil related properties also strategical purposes such as travel or defense related conditions are relevant.

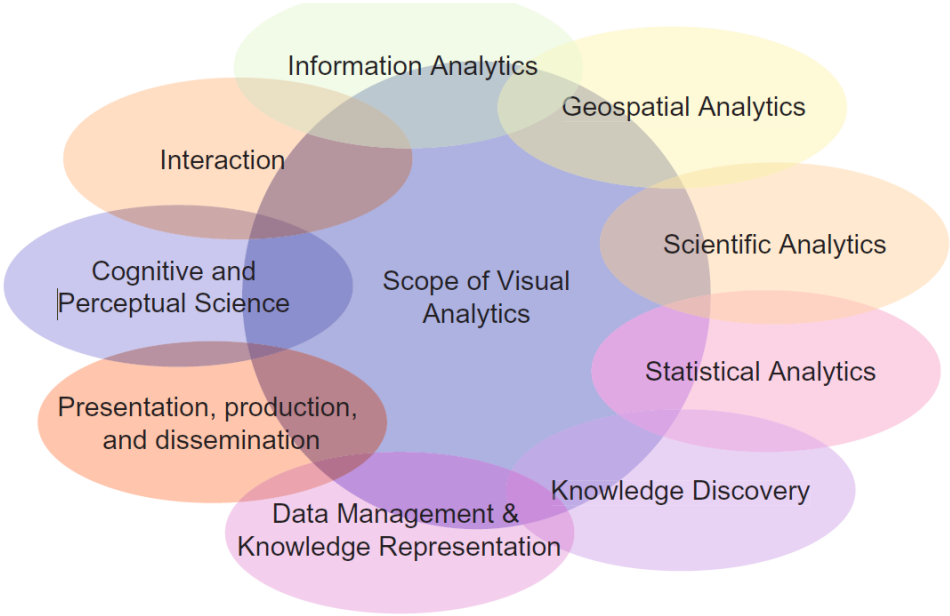

Figure 1. Overview of the related topics of visual analytics (Keim 2008)

To facilitate the analysis for archaeologist, a tool is needed which provides the various environmental data. Additional statistic and spatial analysis tools are required which are specifically tailored for settlement pattern analysis in the Amazon.

3. Requirement Analysis

As mentioned above there is no data source which provides the geographical and thematic information about the excavation sites digitally. Which means, beside the thoughts about tools and additional background data, one important step would be to make such a dataset available for download.

A visual analytics application for archaeological approaches in the Amazon needs certain functionalities. In the context of settlement pattern analysis it means that the application should provide data and tools that factor the environmental structure as well as additional settlement related information. To identify the necessary data the relevant literature was reviewed.

The observed region – Amazonia – is huge why we assume, that there are varying environmental conditions, based on the location and the sociocultural characteristics, that had influence on the settlement behavior. Whereas some of the relevant data, such as soil type, climate or vegetation type, may be important for all cultures, others may depend on the capabilities and needs. To factor that, a dynamic design is required, which allows statistical and spatial analysis based on a user-defined selection of the excavation sites.

3.1. Required Data

The most important data is the information about the known excavation sites and their location. This data should be semantically enriched by the culture(s) and tradition(s), radiocarbon data (if available), the data source and the size (if available). Winkler-Prins & Aldrich (2010) also included some additional information, e.g. the estimated accuracy or if a GPS position was given.

Terra preta sites relate to sedentary historical settlements in Amazonia. (Smith 1980). McMichael et al. (2014) tried to predict pre-Columbian anthropogenic soils (terra pretas) in Amazonia and included 45 environmental layers which are subdivided into bioclimatic, soil, terrain, hydrological and geological topics. They narrowed down the number of variables to 22 by performing cross-correlation analysis. The results show that terrain, geology and hydrology have a greater influence than climate. Elevation had the greatest impact followed from by the geological province and the river-

ine settings. Denevan (1996) presented a bluff model of riverine settlement in which he emphasizes the importance of bluff zones along the major rivers. That underlines the influence of the terrain. It is said that settlements are always close to rivers on the bluff zone to avoid seasonal flooding (Koch-Grünberg 1923). Besides the data mentioned above, also strategic interests are factored. That is due to the assumption that different simultaneous cultural settlements had different functions. This means that e.g. some settlements were located so that the visibility down the river was good and potential enemies could be identified early (often directly at the Amazon River). Another example is the settlements close to the waterfalls which provide certain resources (e.g. stones) which cannot be found further downstream in direction to the Amazon River. That means that these settlements were the only ones which could provide a resource everybody else needed.

In Amazonia there are three main river types (Sioli 1983) which differ in the food supply. According to McMichael et al. (2014) the river types as no influence on the probability of terra preta sites. That is because of the scale and the global approach. The settlement behavior can vary according to the observed culture, and therefore on a smaller scale, why the different river types are included in our research. Especially because the cultures knew about the soil fertility of the alluvial land of the white water rivers (várzea) and the richness of fish in the várzea lakes (Sioli 1983).

To prevent the usability of a visual analytics application, the total number of layers (overlays) is limited. Some of the environmental information is not simply visualized but also precalculated layers are derived based on the base data. The distance measures (e.g. distance between the excavation sites and the rivers) are calculated beforehand and can be used as an additional overlay.

3.2. Required Tools

Besides the usage of different environmental information some spatial and statistic tools would factor the usability of the application.

Depending on the Visual Analytics application, the variety of provided tools differs (Harger & Crossno 2012). Tory and Möller (2004) classified visual analytic tools based on algorithm data types, such as continuous versus discrete values, and subdivided those categories into the number of dimensions and algorithms on graphs and trees. Harger & Crossno (2012) used four categories, namely graphs, trees, n-dimensional data (both continuous and discrete) and geospatial/spatio-temporal visualizations. Each of these categories holds a number of tools to perform visual analytics. Due to the available data and the aim of performing settlement pattern analysis, only

very little of the tools are necessary (e.g. the graph or tree related tools are not needed).

The period of time in which cultures and traditions existed are roughly known. A bounding ellipse around either each culture or each tradition may help to identify a temporal change (e.g. movement towards the Amazon mouth). Even though there is no density information included it also gives a first overview about a spread of the observed group. Additionally a time axis can give an overview about the appearance.

To factor the settlement density, a point density tools is required. This tool calculates the density based on the user-defined selection. Due to this, cultural or traditional differences can easily be identified.

To estimate an area of daily actions a buffering tool is necessary. It is only a broad approximation but gives a first insight. The buffer distance is either a user defined value or a field in the data base to visualize different scenarios. In the latter case, it is possible to factor different radiuses for different former settlement.

Figure 2. Mosaic plot which displays the environmental patternd for three cultures

Another plot is the mosaic plot which helps to identify patterns of intra- and inter-cultural as well as intra- and inter-traditional relations. It shows the variation of environment based on the observed group(s) (see Figure 2).

To get an impression of the surrounding environment a radar chart should visualize the availability. For each excavation site the distance to user defined resources are displayed according to their cardinal directions (see Figure 3).

Figure 3. Radar chart showing the distance to two selected environmental variables

Archaeologists still excavate former settlements in Amazonia. Due to that it is recommended to implement a tool to easily add and change excavation sites. The necessity of changing is because maybe GPS coordinates are accessible

or another, more accurate data source is found. To avoid misuse a login is required. Furthermore a versioned database would help to restore a previous status, if an unwanted change was submitted.

Besides the mentioned tools, also some statistical tools are required. The distance to the rivers (and different types of rivers) will be displayed using a histogram and a boxplot. One can easily see if most of the settlements were close to rivers or certain river types. It is also possible to see if there are known outliers which may lead to new settlement theories. This might happen due to rescue excavations which also lead to unexpected findings in greater distance to rivers. This can include statistical tests against complete spatial randomness.

A contingency table which is generated dynamically is helpful. The environmental data as well as the excavation sites are user-defined. This helps to easily identify important factors (e.g. 80% of all Konduri findings are in 1 kilometer distance to a river)

4. Implemented Web GIS (beta)

The implementation of the described application is still work in progress and is extended continuously. For now only very little things are usable. Available are several layers (besides an aerial image) which can be used as an additional overlay. The excavation sites are displayed and are available for download. Besides the coordinates extra thematical attributes such as the culture and tradition or radiocarbon dates are provided. For now the download format is shapefile but it is planned to also provide a google kml file. The reason is that a lot of people are used to google earth but are not familiar with GIS systems such as Quantum GIS or ArcGIS.

Our Database contains the most accurate source of terra preta sites and the access is free. Besides the location we provide the information about the excavation site name and number to perform further research at the iphan database. Also there is information about the original data source(s).

The project can be seen here:

http://terrapreta.geo.uni-augsburg.de/

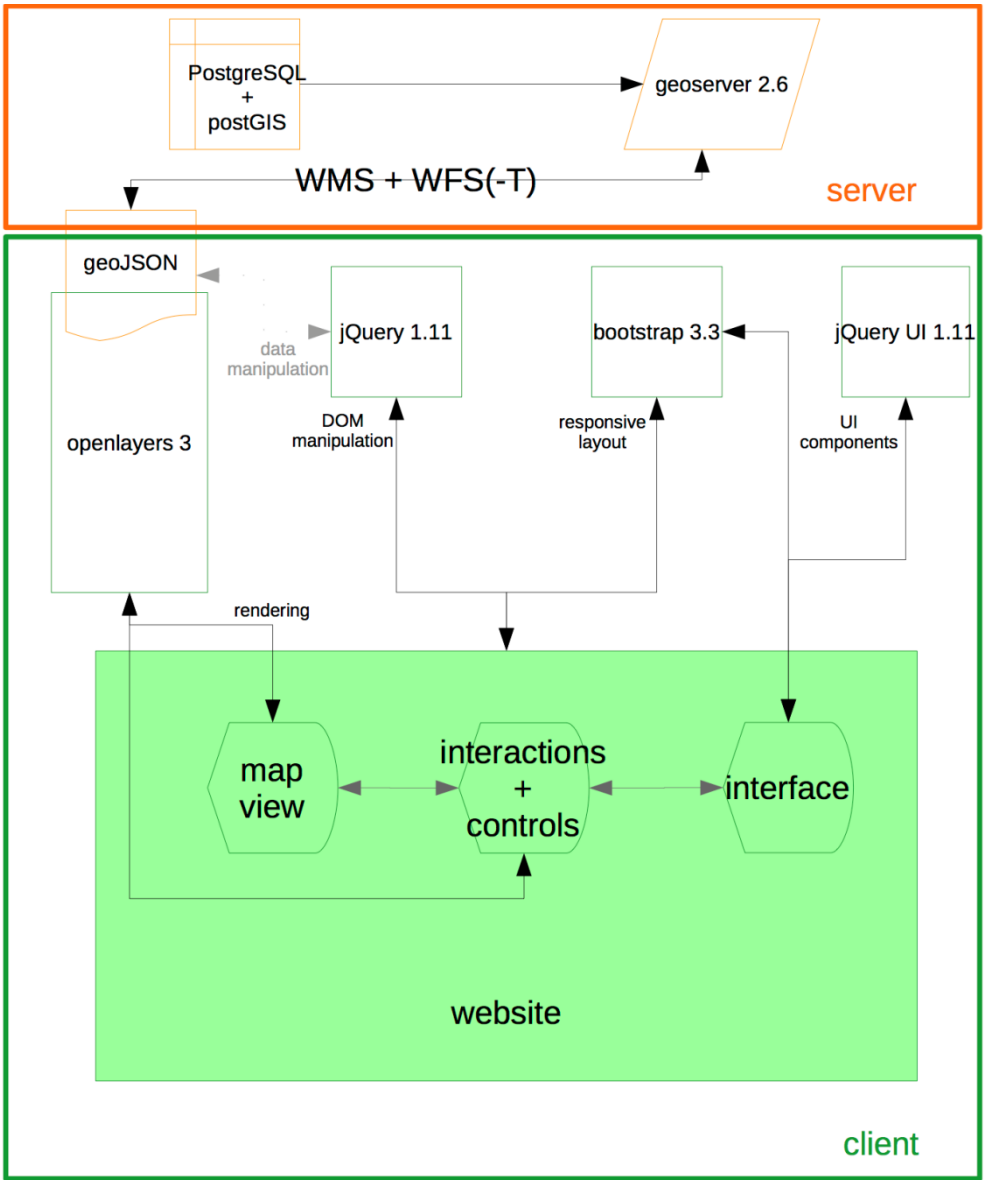

Figure 4. Schematic representation of the WebGIS application

The functionality will be extended continually. But the schematic framework (see Figure 4) is already fully implemented. Without going to much into detail some thing can be mentioned. Underlying is a PostgreSQL database with its spatial extension postgis. This database stores the excavation sites with its thematic attributes as well as the environmental information. The geoserver provides WMS (Web Map Service) and WFS

(Web Feature Service) depending on the requested data (e.g. based on the selection). The listed frameworks, such as jQuery, bootstrap or geoJSON are used for the interaction between client and server and to facilitate the user interaction. The result is a website which can display maps, provide tools and react to the users input.

5. Conclusion

The information about the precolonial cultures in Amazonia is inherently vague and there is a great need to provide a complete as possible database to simplify and facilitate further research. A web GIS is a suitable tool for such a complex issue because it factors the possibility of visual analytic technics and does not need expert knowledge in geoinformatics or GIS. The presented requirement analysis shows, that not the whole spectrum of visual analytic tools is required. A small excerpt, tailored to the needs of settlement pattern analysis seems to be sufficient to perform the task. More important than the toolset itself is the dynamic organization of the tools. That is due to the fact that the settlement patterns may vary across the cultures or traditions. It is not appropriate to except the same behavior in this huge spatial and temporal variation.

For now, the historical environmental data is not factored which is maybe relevant to analyze settlement patterns based on environmental information. It can be assumed – due to the position of the bluff zones – that the course of the river remained more or less the same (with minor variations). The course of the river was manually corrected only in the case of river dams, because otherwise the former settlements would be located in the water. Other important resources, such as wood or stones – were either ubiquitous or are still located at the same spots. That leads to the assumption that only little environmental data actually needs historically enriched information.

The presented requirement analysis focuses on Amazonia which means that other, similar applications maybe need additional or different input data and tools. Countryman et al. (2011) presented a study about predictive modeling of agricultural terraces in Abruzzo, Italy. Due to the topographical differences variables such as the aspect were also factored.

Scale does matter. In this approach only macro scale settlement behavior is displayed. Intra site settlement patterns are not factored. One excavation site is seen as a point geometry and not as a region with several houses etc.

The ability to extend and change the underlying geodata, namely the excavation sites, without GIS, or programming knowledge raises the acceptance of this application.

References

Andrienko N, Andrienko G (2006) Exploratory Analysis of Spatial and Temporal Data, Springer, Heidelberg

Andrienko G, Andrienko N, Demsar U, Dransch D, Fykes J, Fabrikant SI, Jern M, Kraak MJ, Schumann H, Tominski C (2010) Space, time and visual analytics, International Journal of Geographical Information Science 24(10):1577-1600

Badie B, Berg-Schlosser D, Morlino L (2011) International Encyclopedia of Political Science, Sage Publications Inc., Thousand Oaks

Countryman JR, Carrier SC, Kane SE (2012). Predictive Modelling: a Case Study of Agricultural Terraces at Monte Pallano (Abruzzo, Italy), Zhou, M., I. Romanowska, Z. Wu, P. Xu and P. Verhagen (eds.), Revive the Past. Computer Applications and Quantitative Methods in Archaeology (CAA). Proceedings of the 39th International Conference, Beijing, April 12-16, Pallas Publications, Amsterdam:307-317.

Denevan WM (1996) A Bluff Model of Riverine Settlement in Prehistoric Amazonia, Annals of the Association of American Geographers 86(4):654-681

Harger JR, Crossno PJ (2012) Comparison of open-source visual analytics toolkits, Proceedings of the SPIE 8294: 82940E-82940E-10, doi:10.1117/12.911901

Hilbert PP (1968) Archäologische Untersuchungen am mittleren Amazonas: Beiträge zur Vorgeschichte des südamerikanischen Tieflandes, Marburger Studien zur Völkerkunde, Reimer Verlag, Berlin

Jones LV ed. (1986) The collected works of John W. Tukey: Philosophy and principles of data analysis 1965-1986 (Vol. IV), Chapman & Hall, London

Keim DA, Mansmann F, Schneidewind J, Thomas J, Ziegler H (2008) Visual Analytics: Scope and Challenges, Springer, Heidelberg

Koch-Grünberg, T (1923) Vom Roroima zum Orinico Band III, Verlag Strecker und Schröder, Stuttgart

McMichael CH, Palace MW, Bush MB, Braswell B, Hagen S, Neves EG, Silman MR, Tamanaha EK and Czarnecki C (2014) Prediciting pre-Columbian anthropogenic soils in Amazonia, Proceedings of the Royal Society of London 81(1777) DOI:10.1098/rspb.2013.2475.

Mitchell A (1999) The ESRI Guide to GIS Analysis – Volume 1: Geographic Pattern & Relationships, Esri Press, Redlands

O'Sullivan D, Unwin D (2010) Geographic Information Analysis 2nd edition, John Wiley & Sons Inc., New Jersey

Roosevelt AC, Lima da Costa M, Lopes Machado C, Michab C, Mercier N, Valladas H, Feathers J, Barnett W, Imazio da Silveira M, Henderson A, Sliva J, Chernoff B, Reese DS, Holman JA, Toth, N, Schick K (1996) Paleoindian Cave Dwellers in the Amazon: The Peopling of the Americas, Science 272:373-384. doi:10.1126/science.272.5260.373

Sioli H (1983) Amazonien Grundlagen der Ökologie des größten tropischen Waldlandes, WVG, Stuttgart

Smith NJH (1980) Anthrosols and Human Carrying Capacity in Amazonia, Annals of the Association of American Geographers 70(4):553-566

Thomas JJ, Cook, KA, eds (2005) Illuminating the Path: Research and Development Agenda for Visual Analytics. IEEE Press, Los Alamitos

Tory M, Möller T (2004), Rethinking visualization: A high-level taxonomy, Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on:151–158

Tukey JW (1977) Exploratory Data Analysis, Addison-Wesley; Reading

WinklerPrins AMGA, Aldrich SP (2010) Locating Amazonian dark earths: creating an interactive GISof known locations. Journal Latin American Geography 9:33–50 doi:10.1353/lag.2010.0029