Analysing settlement patterns in the Bergisches Land, Germany

Irmela HERZOG
The Rhineland Commission for Archaeological Monuments and Sites, Bonn, Germany

Abstract: The nearly complete list of Medieval settlements in a study area of the Bergisches Land, Germany, and a historical map of 1715 form the basis for analysing the settlement pattern of Medieval and early modern farms in this rural region. Only few archaeological remains of pre-Medieval periods have been recorded in this hilly study area with lots of small streams and at best moderately fertile soils, so most probably the population and the number of farms increased considerably during the Middle Ages. Settlement sizes (e.g. one, two, or three farms, hamlets, or villages with a church) are shown on the early 18th century map depicting nearly all Medieval settlements known from historical sources as well as a small portion of additional settlements. This study tries to retrodict likely settlement locations in this rural area, i.e. to identify the main factors governing the location of farms and villages. This is an unusual task due to the dense distribution of these locations. Several possible factors for settlement location are tested, and a retrodictive model combining relevant variables is presented. The analysis indicates that some variables with strong retrodictive power do not apply for single farmsteads. The landscape attributes identifying possible settlement locations tell only part of the story. Additional analysis is required to identify the distribution of the settlements with respect to their neighbours, such as regular distances between settlements of similar sizes as proposed by Christaller’s Central Place Theory. Least-cost Thiessen territories are calculated for all settlements on the historical map of 1715. It is shown that the Thiessen territory size depends on the settlement size and this indicates that hardly any space is left for additional settlements.

Keywords: settlement patterns; retrodiction; landscape attributes for settlement location; Kolmogorov-Smirnov test; Thiessen polygons

Introduction
The initial aim of this study was to reconstruct the settlement spreading processes in the study area in Medieval times. The study area in a hilly region east of Cologne was sparsely populated before Medieval times because the other regions of the Rhineland offered more attractive natural conditions for farming. From the 9th century onward, nearly all the fertile regions of the Rhineland were occupied by farms, and at the same time the climatic conditions for agriculture improved, so that the population increased in the study area (LANDSCHAFTSVERBAND WESTFALEN LIPPE & LANDSCHAFTSVERBAND RHEINLAND 2007, on CD: 282). The historical data for this area includes a list of place names and the year when these sites were first mentioned (PAMPUS 1998). Nearly all of the place names could be located on maps created in the late 19th century, and this allowed georeferencing the centres of the settlements (Fig. 1). Only very few of the Medieval settlement locations were abandoned. 54% of today’s settlements in the study area were mentioned in historical sources dating back to years earlier than 1500 AD. 82% of today’s
settlement locations were mentioned in historical sources before 1600 AD. So it seems that this region with fairly bad soils saw a development from empty to full within a few hundred years.

Fig. 1 – The study area east of Cologne, Germany, covers nearly 1000 km² of the hilly region Bergisches Land. A large number of small rivers and creeks can be found in this area. Most parts of the study area are in an altitude range between 200 and 400 m asl. In the background of the settlement distribution map a digital elevation model (DEM) is shown. This DEM with a resolution of 25 m was provided by the Ordnance Survey Institution responsible for this area (geoBasis NRW).

Initially, it was planned to model the spreading of settlers that move into unknown terrain. But the high density of settlement locations shown in Fig. 1 indicates that adult children seeking a new place to live knew the surrounding area and probably found an appropriate settlement location in the vicinity. For this reason the focus of the study was shifted towards an issue that was only one step on the agenda in the first place: The aim of the present study is to identify possible locations of Medieval and early modern farms. The question is if settlement locations were chosen at random within this study area of high settlement density or if landscape attributes for settlement location can be identified. Another issue are the settlement sizes. Previous studies of settlement patterns in this area (HERZOG 2012; 2014) had attributed equal weight to each settlement location claiming that nearly all settlements were more or less of equal small size in Medieval times, but the map created by Erich Philip Ploennies in the years up to 1715 AD shows different symbols to indicate different settlement sizes. The mathematician Erich Philip Ploennies worked as a cartographer for the rulers of the Bergisches Land. Unfortunately, part of the study area belonged to another
territory at that time so that it is not covered by Ploennies’ maps (Fig. 2). The western part shown in Fig. 2 has an area of 412 km² and includes 749 settlements known in 1715 AD. Within the eastern part covering 241.6 km² 254 settlements can be found.

Ploennies also gives a short description of the landscape and the main jobs of the people living there. Although two towns with walls existed and some quarry, mining and metal working activities are recorded it seems that even the people in the towns mostly lived on farming. Mainly oat was grown because this crop does not require good soils.

Identifying landscape attributes for settlement location

This section deals with identifying landscape attributes for Medieval and early modern settlement location in the study area. This is closely related to indicative predictive modelling used in archaeology to “predict the probability of archaeological settlements occurring in unsampled landscapes” (CONOLLY & LAKE 2006: 179-186). The Ploennies map provides a complete sample, and therefore the aim is retrodiction rather than
prediction. In the previous section soil quality was identified as one possible relevant landscape attribute by visually inspecting the soil quality map. Box plots (also known as box-whisker plots; DE SMITH et al. 2007: 182-183) provide an easy way to compare the distribution of the settlement’s soil quality with the overall distribution of the soil quality within the study area. The overall distribution is found by subdividing the study area into cells of 250 m x 250 m and calculating the average soil quality within each cell. This process results in 16,642 cells, and the first box plot in Fig. 3 shows the distribution of the soil quality for the grid cell centres. The most important key figure in the box plot is the median, that is the middle value in the sorted list of values so that 50 % of the values are below and 50 % above the median. The median of the overall distribution (41.3) is marked by a dotted orange line. The median of the soil quality of all settlements considered is above this value (43.5). The box plot consists of a box and of whiskers. The median is represented by a horizontal line within the box, the lower and upper limit of the box mark the quartiles: 25 % of all values are below the lower quartile and 25 % of all values are above the upper quartile. For the overall distribution of soil qualities, the lower quartile is 36.9, and the upper quartile is 44.7. So soils with a quality within the range between 36.9 and 44.7 cover 50 % of the study area. Moreover, the whiskers of the box plots indicate the range of the values. In some box plots the maximum value is cut off due to space limitations. The box plots in Fig. 3 support the hypothesis that many single farms are located on poor soils whereas all other settlement types prefer good soils.

Fig. 3 – The box plots show that the Medieval settlements in the study area prefer good soils. However, many single farms occupy areas with lower soil quality than the median quality within this region.

In their chapter on predictive modelling, CONOLLY and LAKE (2006: 182) recommend identifying relevant landscape attributes by univariate statistical tests. Kolmogorov-Smirnov tests are appropriate for comparing the distribution of the grid centre soil quality values with that of the settlements. The free statistics software PAST (HAMMER 2014) applied for testing supports only two-sample two-sided tests. Although one-sample one-sided tests might be more suitable, the PAST results are valid when rejecting the null hypothesis because the probabilities resulting from a one-sample one-sided test are lower than that of a two-sample two-sided test. Another issue is the fact that the test requires independent observations, and with spatial data there is always some autocorrelation. It seems that the impact of this issue is limited because the autocorrelation of the settlement locations is lower than that of random locations in most test settings considered in this study. Details concerning the impact of autocorrelation on test outcomes are given in HERZOG (2014). The result of a Kolmogorov-Smirnov test in PAST is the probability p that the values of two samples are drawn from the same distribution. The PAST result for comparing the distribution of the grid
centre soil values with that of all settlements considered (n = 1312) is 6.259E-38, the comparison with all Medieval settlements (n = 789) results in a p-value of 2.18E-33. For all subsets of the settlements analysed in Fig. 3 the p-values calculated by PAST are low, the highest probabilities are for single farms (n=395, p=1.794E-05) and hamlets, villages, towns (n=146, p=8.807E-05).

Other landscape attributes for settlement location typically analysed in predictive modelling are: slope, aspect, distance to the nearest stream, and relief (CONOLLY & LAKE 2006:180-182; EJSTRUD 2003). Both the accuracy and resolution of the digital elevation model (DEM) as well as the method used for deriving the variables from the DEM have some impact on the outcome. The modern DEM used in this study has a resolution of 25 m x 25 m, is based on LiDAR data and was provided by the Ordnance Survey Institution responsible for this area (geoBasis NRW). Since Medieval times human activities like the construction of dams for water reservoirs, motorways and quarries, as well as natural processes like erosion, meandering of rivers and flooding changed the surface of the terrain within the study area. But reconstructing the Medieval terrain is beyond the scope of this study. Slope and aspect were calculated using Vertical Mapper, a plugin for MapInfo (DE SMITH et al. 2007: 260).

The box plots in the upper row of Fig. 4 provide evidence that all settlements considered prefer low slopes irrespective of period or size of the settlement. The Kolmogorov-Smirnov p-values are even lower than for soil quality (p=2.307E-55 for all settlements and p=4.731E-39 for Medieval settlements) indicating that there is a high probability that the settlement locations were not chosen at random with respect to slope.
Farming settlements consist of houses surrounded by fields or meadows (NICKE 1995: 52), so not only the slope of the house location but also of a catchment area for each settlement might be relevant. Therefore the mean slope within a catchment area covering 525 m x 525 m (6.25 hectare) was calculated for each cell. EJSTRUD (2003) uses the term relief for a similar key figure, i.e. the mean slope within 500 m. Comparing the leftmost box plot in the lower row of Fig. 4 to all the other box plots in this row results in the conclusion that local mean slope is not as relevant as slope for settlement location. The p-values of the Kolmogorov-Smirnov test are higher compared to those resulting from considering slope at the settlement location only (p= 5.806E-34 for all settlements and p=5.913E-29 for Medieval settlements). Testing the outcome of different catchment sizes remains a task for future work. The ideal catchment size depends on the size of the gardens, ploughland, and meadows used by each farmstead. According to Henning (1994: 155-159), about 7 to 8 hectare are necessary for supplying a family with crops, and until the middle of the 14th century, the farmland of most farms did not exceed 8 hectare. Due to the fairly bad soils in the study area, the farmland used by a family was probably larger.

Box plots analysing the settlement locations with respect to aspect and local average aspect (within a catchment area of 525 m x 525 m) were also created but are not shown here. According to these box plots the aspect values of the Medieval and early modern settlements do not differ much from that of all grid cells except that grid cells facing west do not include as many settlements as expected for a random distribution. The p-values of the Kolmogorov-Smirnov test with respect to aspect are still very low, but a lot higher than for soil quality and slope (p=3.433E-10 for all settlements and p=3.886E-08 for Medieval settlements). The p-values for local average aspect are even higher: p= 0.09438 for all settlements and p= 0.3159 for Medieval settlements.

The upper row of Fig. 5 shows that Medieval settlements mentioned before 1400 AD tend to be located in low altitudes benefiting from a somewhat longer annual time span for vegetation growth. But the p-values of the Kolmogorov-Smirnov test with respect to altitude significantly exceed those of soil quality and slope (p=6.961E-07 for all settlements and p=2.594E-08 for Medieval settlements). Again many single farms can be found in areas with less favourable conditions for farming.
Fig. 5 – The box plots in the top row show that the Medieval settlements first mentioned before 1400 AD prefer low altitudes. Whereas single farms tend to be located on higher altitudes than the median, the sites consisting of two to four farms can be mostly found on altitudes below the median. The box plots in the bottom row provide evidence for a low local prominence of the settlements considered.

The local topographic prominence (with respect to elevation) of the settlements was calculated based on ideas proposed by LLOBERA (2003): Whereas Llobera subtracts the local mean value, the local median value was subtracted in this case in order to reduce the impact of outliers. Using the plugin Sextante (option: focal statistics, Median (neighbourhood)) of gvSIG, the local median value was determined within a radius of 200 m. Topographic prominence plays a role for settlement location in wet areas where small artificial or natural hills usually offer a dry and safe place for houses. But in the region considered here a different type of settlement location is to be expected: According to NICKE (1995: 49), early settlements in this region can mostly be found in areas of natural depressions with sources (in German: Quellmulden). Such a location does not exhibit local prominence but provides some shelter. The bottom row of Fig. 5 shows that the settlements considered mostly prefer locations with low local prominence (p=8.697E-53 for all settlements and p=6.393E-34 for Medieval settlements).

Distance to streams is another landscape attribute considered in many predictive modelling studies. In the study region with many steep slopes, map distance does not seem appropriate for estimating the effort of covering distances in Medieval times. Instead least-cost distances were calculated. The cost model was derived from Medieval trade routes recorded in this area (NICKE 2001). Least-cost path calculations applying a cost model avoiding wet soils and steep route segments with a gradient above 13 % produced least-cost paths that fitted best to the trade routes (for details see: HERZOG 2013a).

For calculating the distance to streams a vector layer with lines was available indicating also the breadth of the streams. First the least-cost distance to all (possibly navigable) streams with a breadth of 3 m or more was investigated. The box plots (not shown) indicated that a low least-cost distance to these streams is not
important for the settlements considered. The least-cost distance to all streams with a breadth of up to 3 m in the vector layer has some retrodictive power (top row in Fig. 7), with $p=3.601\times10^{-32}$ for all settlements and $p=3.18\times10^{-24}$ for Medieval settlements. But looking at the settlement locations on historic maps created the impression that the stream vector layer did not contain all relevant creeks.

Fig. 6 – The map on the left shows the accessibility to the nearest stream with a breadth of up to 3 m based on a vector layer. This map does not take small creeks into account, which can be identified by flow accumulation calculations. An accessibility map with respect to cells with high flow accumulation is shown on the right.

According to Nicke's description of favourable settlement locations, settlements are to be expected near sources of small creeks, and many of these creeks are missing in the vector layer. Therefore a flow accumulation layer was derived from the DEM using gvSIG Sextante (options: Basic hydrological analysis, Flow accumulation, Multiple Flow Direction). This layer includes nearly all cells passed by the lines in the vector layer and indicates the location of many small creeks in the study area.

Fig. 6 shows accessibility maps with least-cost distances to vector layer streams with a breadth of less than 3 m (left) and to flow accumulation cells (right), illustrating the large difference between the vector layer and the multitude of creeks derived by flow accumulation. The least-cost distance to flow accumulation cells has more retrodictive power than distances calculated on the basis of the vector layer only (Fig. 7), with $p=3.111\times10^{-90}$ for all settlements and $p=3.933\times10^{-63}$ for Medieval settlements.
Fig. 7 – The box plots in the top row show that most settlement types considered prefer high accessibility to the nearest creek or small river. The box plots in the bottom row indicate that if small creeks are taken into account (i.e. the calculation is based on flow accumulation), the importance of a short distance to surface water is more evident.

Another landscape attribute considered in this study is accessibility in terms of movement (Herzog 2013b) applying the cost model derived from the trade routes. But this variable turned out to be of lower impact on the settlement location than soil quality, local prominence or slope, with p=3.998E-05 for all settlements and p=0.004769 for Medieval settlements. This may be due to the fact that the settlements tend to be close to creeks, and according to the cost model, wet areas are assigned a cost multiplier of 5.

This section presented an analysis of landscape attributes with possible retrodictive power. Four strong impact variables could be identified: Soil quality, slope, local prominence and least-cost distance to flow accumulation. Variables with moderate retrodictive power are local average slope and aspect. Except for soil quality, these variables are important for settlement locations of all sizes and all time intervals considered. Single farm locations are different from settlements mentioned before 1400 AD with respect to two landscape attributes, namely altitude and soil quality.

**A retrodictive model based on strong impact variables**

Based on the results presented in the previous section, a simple retrodictive model was created relying only on the strong impact variables. Cut-off values were selected according to a 90 % rule: The slopes of 90 % of the settlement locations are below 18%, only 10 % of the settlements are situated in areas with a soil quality below 38.5, and the least-cost distance to flow accumulation cells is below 9 for 90 % of the settlements. On the basis of the histogram for the settlement's local prominence distribution, the 90 % range consists of all values in the range between -13.75 and 6.00.
Applying the 90 % rule, the model consists of all 250 m x 250 m grid cells within the study area for which the slope is below 18 %, the average soil quality exceeds 38.5, the least-cost distance to flow accumulation is below 9 and the local prominence value is in the range between -13.75 and 6.00. These conditions are fulfilled by 5474 (32.9 %) cells within the study area (Fig. 8).

The resulting map consists of a large number of small patches i.e. areas with high probability of settlement locations. The grid cells within the model include 558 (47.9 %) of the settlements considered. A performance indicator for predictive models was proposed by Kvamme (referred to by EJSTRUD 2003; CONOLLY & LAKE 2006: 185) and is known as Gain: Gain = 1 - %area / %sites where %area is the percentage of green cells in Fig. 8, and %sites is the percentage of sites within green cells. For all settlements considered the Gain is 0.41; the Gain for settlements mentioned before 1400 AD is somewhat higher: 0.48. This is well within the performance range of the best methods tested by EJSTRUD (2003). The Gain achieved with the retrodictive model is not impressive, but this result shows that in this study area with a high density of settlement locations, these locations are not chosen purely at random. A better result in terms of Gain may already be achieved by a more sophisticated choice of cut-off values or by including additional impact variables. Applying a more refined statistical method for creating the retrodictive model (e.g. the approaches discussed by EJSTRUD (2003)) is likely to improve the performance as well. Moreover, a separate retrodictive model for single farms might be appropriate: The single farms are mainly located in the north-western part of the areas covered by Ploennies. The single farm “hot spot” area was delineated by kernel density estimation (CONOLLY & LAKE 2006: 175-179), this area covers 13.4 % of the
Ploennies areas and includes 44.1 % of the single farms, i.e. the Gain value is 0.696. Unfortunately, this technique cannot be used to predict the single farm locations within the area not covered by Ploennies.

**Settlement patterns beyond landscape attributes**

The landscape attributes identifying possible settlement locations tell only part of the story. A supplemental analysis deals with the distribution of the settlements with respect to their neighbours. As mentioned above, each of the rural settlement locations was surrounded by gardens, ploughland, meadows and possibly forest areas (NICKE 1995: 52) and this should have some impact on the settlement pattern.

The settlement distribution map in Fig. 2 shows that there are no regular distances between settlements of similar sizes as proposed by Christaller’s Central Place Theory (RENFREW & BAHN 1996: 170-171). As shown above, the single farm settlement locations are not regularly distributed within the study area. Christaller’s theory is based on the assumption that settlements are located in a uniform landscape where topographic features like mountains and variations in soil quality are absent. The analysis presented in the previous section shows that this assumption does not hold for the study area considered.

RENFREW & BAHN (1996: 171) propose calculating Thiessen polygons (also referred to as Voronoi diagram or Dirichlet tessellation; CONOLLY & LAKE 2006: 211-212) as an alternative to Central Place Theory. By applying Thiessen polygons, a territory is allocated to each settlement. Usually this allocation process is based on map distances but the method can be readily adjusted to support least-cost distances (HERZOG 2013a). Fig. 9 presents a thematic map based on the sizes of the least-cost Thiessen patches calculated for the Medieval and early modern settlements within the areas covered by Ploennies’ maps.

This map clearly shows different settlement patterns in the western and eastern parts of the study area. The territories apportioned to the settlements in the eastern area are in general bigger than those in the western part. According to Fig. 2 most single farms can be found in the north-western part. This observation suggests that the territory size depends on the settlement size. Therefore, the least-cost Thiessen patch sizes were analyzed with respect to the Ploennies settlement sizes (Fig. 10) omitting least-cost Thiessen patches at the border of the study areas because these patches are arbitrarily cut off at the limits of the study area.

The box plots in Fig. 10 indicate that the Thiessen patch size depends on the settlement size. The Thiessen patches of single farms are larger than expected when compared to those of settlement locations with two or three farms. But many single farms can be found on bad soils and on high altitudes with a shorter annual vegetation growth time span. Such single farms require a larger area per person than settlements in a more favourable setting for agriculture. The median of the single farm patch sizes (44.86 hectare) substantially exceeds the figure of 8 hectares given by HENNING (1994: 159) for a typical family-run farm before the middle of the 14th century. The median size of the territories allocated to settlements consisting of two farms is only somewhat higher (45.98 hectare).
Fig. 9 – Least-cost Thiessen patches for all settlements considered with colours indicating the size of the patches. For analysing the sizes, only the patches within the areas delimited by the brown borders are considered.
Both concepts, Central Place Theory and Thiessen polygons, are models for a full landscape, with no room for additional settlements. RENFREW and BAHN (1996: 171) point out that Thiessen polygons should only be created for settlements of equal rank because the method does not take the importance of the settlement into account. The results in Fig. 10 indicate that there are exceptions to this rule. The relationship between Thiessen patch sizes and settlement sizes supports the hypothesis that hardly any empty patches were left. Another way to analyse the impact of the settlement size on the settlement pattern is to create a least-cost triangulation (HERZOG 2013a) connecting each settlement by least-cost paths to its neighbours. Only those connections between settlement pairs are considered where at least one settlement has a non-border patch. The analysis of the average path cost between two neighbours shows that the least-cost distance depends on the size of the two settlements considered: For instance, the average path cost between a single farm and a neighbouring single farm is lower than that of the single farm to a neighbouring two- or three-farm settlement, and the average distance to a four- or more-farm location is even bigger. Similarly the average path cost between a four- or more-farm settlement and a neighbouring hamlet is lower than that to a village.

**Future work and discussion**

Most of the issues with archaeological predictive modelling are not relevant for the data set presented in this study: There is no bias in the data set because it is a not a sample but the complete record, the difficulties in interpreting survey data are avoided, and all settlements in the data set were present at the same time. But the high density of settlement locations in the study area caused some worries concerning the successful application of the techniques used in predictive modelling. If nearly all possible patches for farm locations are occupied, newcomers searching for a patch do not have much choice. Nevertheless, some impact variables

![Box plots showing the sizes of the least-cost Thiessen patches](image)
for settlement locations could be identified although this model is far from perfect. These variables are a subset of the landscape attributes typically applied in predictive modelling. Some of the variables tested are mean values for a given catchment size, and the catchment size resulting in the best retrodiction outcome should be identified. As pointed out above, a separate retrodictive model for single farms might be appropriate. Moreover, the interrelationship between the variables should be analysed and taken into account. More refined modelling techniques probably result in higher Gain values.

As an alternative to the inductive approach presented above, a deductive retrodictive model could be generated from Nicke’s Quellmulden hypothesis.

Future work also includes some additional analysis of the distances between neighbouring settlements and the patch sizes. Least-cost distances derived from old trade routes were applied for the calculations above, but a different way to measure distances might be more appropriate. In Fig. 10 the settlements corresponding to the upper whiskers (above the upper quartile) indicate empty patches in the settlement distribution. Future research could investigate why these patches were left empty and if they were occupied in a later time period.

Ideally the territories should be allocated in a different way, blending the approaches of Thiessen polygons and the IDRISI hinterland procedure (EASTMAN 2003: 96): An iterative least-cost spreading process starting at a settlement centre should stop if the settlement’s demand for farmland is satisfied or if all neighbouring cells are already allocated to another settlement. For such a model additional data on the past agriculture in this area is required. It may also be possible to deduce this data from the settlement locations by assessing the impact of aspect, slope, soil quality and streams on the farmland catchment size required for one family. The Ploennies maps could also provide the basis for additional research. For instance, the settlement sizes can be used to estimate the number of inhabitants in the year 1715 AD. Due to the irregular distribution of settlements of different sizes (Fig. 2) within the study area, the number of settlements is no adequate proxy for the number of inhabitants. Moreover, the Thiessen patches could be modified so that forest areas delimited in the Ploennies map are not included.

Initially, a simulation study reconstructing the dispersal of the settlements in this area was planned. For such a simulation a larger connected area covered by Ploennies maps would provide a more appropriate basis.

References


Imprint:
Vienna 2015
http://www.chnt.at/proceedings-chnt-19/
ISBN 978-3-200-04167-7
Editor/Publisher: Museen der Stadt Wien – Stadtarchäologie
Editorial Team: Wolfgang Börner, Susanne Uhlirz
The editor’s office is not responsible for the linguistic correctness of the manuscripts.
Authors are responsible for the contents and copyrights of the illustrations/photographs.