Testing Models for Medieval Settlement Location

Irmela Herzog

Abstract This contribution investigates two models for the spread of Medieval settlements in the landscape known as Bergisches Land in Germany. According to the first model, the spread was closely connected with the ancient trade routes on the ridges. The alternative model assumes that the settlements primarily developed in the fertile valleys. The models are tested in a study area for which the years are known when the small hamlets and villages were first mentioned in historical sources. It does not seem appropriate to apply straight-line distances in this context because the trade routes of that time include curves. Instead an adjusted distance metric is derived from the ancient trade routes. This metric is applied to generate a digital raster map so that each raster cell value corresponds to the adjusted distance to the nearest trade route (or fertile valley respectively). Finally, for each model a Kolmogorov–Smirnov test is applied to compare the adjusted distances of the Medieval settlements with the reference distribution derived from the appropriate raster map.

1 Introduction

Two models have been proposed for the spread of Medieval settlements in the landscape known as Bergisches Land in Germany. Some experts think that the spread was closely connected with the ancient trade routes which were already in use before the population increase in Medieval times (e.g. Nicke 1995). An alternative hypothesis assumes that the settlements primarily developed in the valleys with good soil (Kolodziej 2005). This contribution investigates the two

I. Herzog (✉)
The Rhineland Commission for Archaeological Monuments and Sites,
The Rhineland Regional Council, Endenicher Str. 133, 53115 Bonn, Germany
e-mail: i.herzog@lvr.de

M. Spiliopoulou et al. (eds.), *Data Analysis, Machine Learning and Knowledge Discovery*, Studies in Classification, Data Analysis, and Knowledge Organization,
DOI 10.1007/978-3-319-01595-8_38,

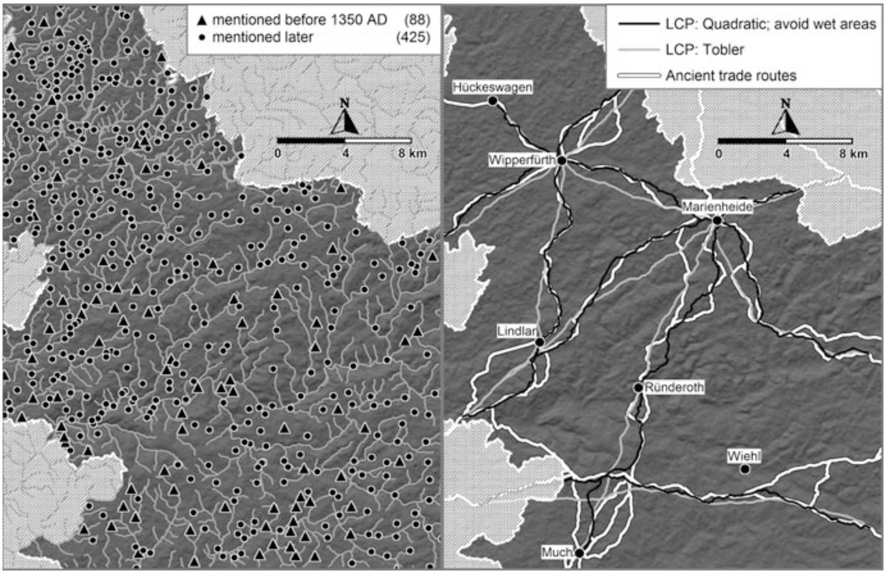

Fig. 1 *Left*: Medieval settlements in the study area, the background shows the terrain and the streams. *Right*: Ancient trade routes and LCPs

hypotheses focusing on an area covering 675 km^2 of the Bergisches Land (Fig. 1). For this study area, a publication is available (Pampus 1998) listing the years when the small hamlets and villages were first mentioned in historical sources, with a total of 513 locations mentioned in Medieval times (i.e. between 950 and 1500 AD), 88 of these were recorded before 1350 AD (early settlements are indicated by triangles in Fig. 1, left). This list of settlements is probably very close to a complete sample, because it includes a high proportion of very small places and only a small fraction of the place names in the list of Pampus could not be located on historical or modern maps. Merely a small amount of the settlements on the list were abandoned. Therefore the assumption is plausible that the sample is not seriously biased. However, a settlement might have been mentioned a long time after its first house had been built. So the history of settlement in this area is perhaps not reflected adequately in the years when the place names were first mentioned.

On the oldest available reliable maps from the study area, drawn in the middle of the nineteenth century, nearly all settlements are still very small, and for this reason it was quite easy to identify the centre of each settlement, which was marked by a dot in a geographical information system (GIS). To test the hypotheses, the distribution of these dots with respect to the targets (i.e. fertile valleys and known ancient trade routes) is compared with the background distribution. In practice, a raster map is created and for each raster cell the distances to the nearest targets are calculated.

The distance distribution of those raster cells which include the settlement dots is compared to the distribution of all raster cells in the study area.

2 Least-Cost Distance

However, the straight-line distance does not seem appropriate for testing the models, because in a hilly terrain with many creeks and small rivers, people did not walk "as the crow flies". The ancient trade routes preserve the movement patterns of the time when they developed. For the study area, these trade routes are described by Nicke (2001) who published overview maps and lists the villages along the routes. So it was possible to outline the routes roughly in a GIS. The route sections connecting the villages were digitized from maps created in the middle of the nineteenth century (Fig. 1, right), assuming that the ancient route layout is still preserved at that time. On the basis of a successful reconstruction of these trade routes by least-cost paths (LCPs), an adjusted distance measure can be derived. Different cost models result in different LCPs. A popular cost model (Tobler 1993; Herzog 2010) depends only on the slope of the terrain and estimates the walking time required:

$$\mathbf{cost}(s) = 10/\mathbf{e}^{-3.5|s+0.05|} \tag{1}$$

where s is the slope (calculated by vertical change divided by horizontal change). The slope is derived from the ASTER digital elevation model (ASTER GDEM is a product of METI and NASA). The resolution of the elevation data is about 30 m.

The goodness of fit of LCPs to the ancient trade routes was assessed by visual impression. Formal methods like that presented by Goodchild and Hunter (1997) were not applied because the improvements after adjusting the model parameters as described below were quite evident. A slope-dependent cost function for wheeled vehicles produced better results than the Tobler cost function. The movement patterns of pedestrians differ from that of carts or wagons with respect to the grade which they can climb efficiently without switchbacks. The term critical slope is used for the transition grade where it is not longer efficient to mount the grade directly. Whereas the critical slope for walkers is at about 25 % (Minetti 1995), a lower critical slope in the range between 8 and 16 % is appropriate for carts or wagons (e.g. Grewe 2004 referring to Roman roads). A simple quadratic function can be constructed with a given critical slope c (Llobera and Sluckin 2007; Herzog 2010):

$$\mathbf{cost}(s) = 1 + (s/c)^2 \tag{2}$$

where c and s are percent slope values. The LCPs resulting from a quadratic cost function with a critical slope of 13 % performed best. However the LCP results came closer to the known trade routes when the slope costs were combined with penalties for crossing wet areas. Initially modern data on streams were used, but some small creeks are not included in the stream data set, and in some cases, the river or creek changed since Medieval times due to meandering or modern construction work. So instead of the modern stream layer, the digital soil map (provided by the Geologischer Dienst, Krefeld, Germany) formed the basis for identifying wet areas. Most small creeks missing in the stream layer but depicted in the mid nineteenth century map are within wet areas indicated on the soil map. Moreover, wet soils are still present in areas where meanders of rivers like the Wupper were located in former times. The LCPs resulting from the penalty factor of 5 for the wet soils coincided better with the trade routes than the paths generated with factors 4 or 6. The best LCP results were achieved when some ford locations identified on historical maps were assigned a lower factor of 2 within the wet areas.

These LCPs are fairly close to the Medieval trade routes (Fig. 1, right). The fit of the LCPs to the trade routes is not perfect which may be caused by several factors, e.g. the inaccuracies of the digital elevation model, landscape change due to modern activities like quarries, mining and the creation of water reservoirs and the fact that additional aspects (beyond slope and avoiding wet areas) might have played a role as well.

So a new metric was defined which takes the slope, the wet areas, and the fords into account. The metric is symmetric because the slope-dependent quadratic cost function is symmetric. The Tobler cost function is not symmetric, but assuming that the same route was chosen for moving to a given target and back, the symmetric function $\text{newcost}(s) = \text{cost}(s) + \text{cost}(-s)$ should be used instead of the asymmetric function $\text{cost}(s)$ (Herzog 2010).

3 Accessibility Maps

On the basis of the distance metric described above, raster accessibility maps are created. The accessibility map with respect to the ancient routes relies on the main routes described by Nicke (2001), ignoring alternative (mostly later) route sections which are also shown in Fig. 1 (right). The digital soil map was used to identify fertile valleys: the soil map attributes include the upper limit of soil quality which indicates agrarian productivity with a maximum value of 100 in Germany, but not exceeding 75 in the Bergisches Land. A thematic map showing this attribute allowed to identify streams surrounded by fertile valleys. So two accessibility maps with two sets of line-shaped targets are created: the first set consists of Medieval trade routes, the second set is formed by streams surrounded by fertile valleys. For each raster cell of the accessibility map, the accessibility value corresponds to the least-cost distance to the nearest trade route or valley with good soils respectively (Fig. 2).

All raster cell values within the study area form the reference probability distribution, and the sample of the cells including the settlement dots will be compared with this reference distribution.

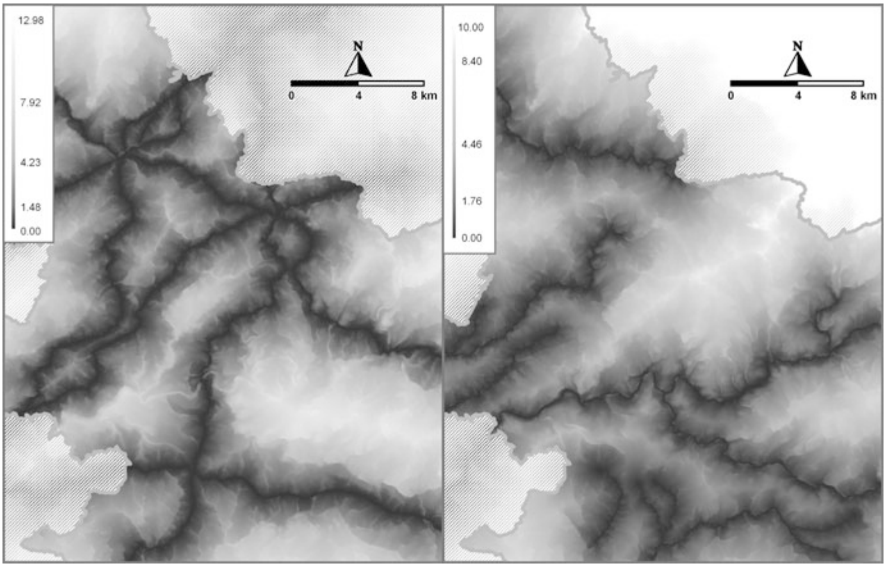

Fig. 2 Accessibility maps, with respect to ancient trade routes (*left*), and fertile valleys (*right*). One distance unit in the map corresponds to the costs of covering 1 km on dry level ground. The legend indicates the quartiles of the distance distribution

4 Testing

One-sample Kolmogorov–Smirnov tests (e.g. Conover 1971) are applied to test the hypothesis that the Medieval settlements are closer (in terms of the least-cost distance) to the linear targets than the reference distribution derived from the appropriate accessibility map. Both reference distributions are skewed to the right so that the popular t-test, which assumes normal distribution, is not applicable. The Kolmogorov–Smirnov test is a non-parametric goodness-of-fit test, which compares the empirical distribution function F_n of the independent, identical distributed (iid) sample consisting of n observations with the cumulative distribution function $F(x)$. The test considers only the maximum vertical distance between these two functions:

$$D = \max_x \; F_n(x) - F(x) \tag{3}$$

The null hypothesis of this one-sided test is rejected if D exceeds a threshold depending on n and the confidence level α. Tables listing the exact threshold values up to $n = 40$ are published (e.g. Conover 1971), for $n > 40$ conservative approximations are used. This allows presenting the test in a diagram, which shows the two distribution functions and their maximum distance (Fig. 3).

Moreover, with a given threshold an upper confidence band for the empirical distribution function can be constructed. The null hypothesis of the goodness-of-fit

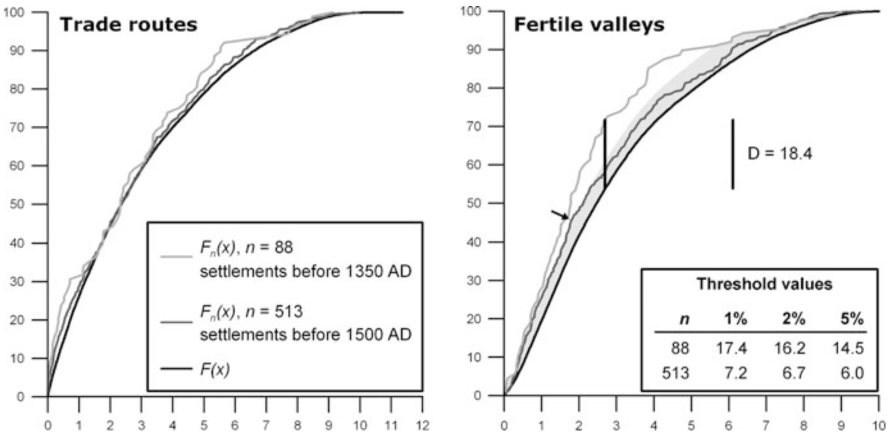

Fig. 3 Kolmogorov–Smirnov tests (one-sided, one-sample). Fertile valleys (*right*): The *black bar* indicates the maximum distance between F_{88} and F; the grey confidence band is drawn for $\alpha = 1\,\%$ and $n = 513$; F_{513} exceeds the confidence band at the position indicated by the *arrow*. On the x-axis the same distance units are used as in Fig. 2

test is rejected if at some point x, $F_n(x)$ exceeds the upper confidence band. So this test is easily intuitive for archaeologists without sound statistical background. Figure 3 (left) shows that the Medieval settlements within the study area are somewhat closer to the trade routes than the raster cells forming the reference distribution, however, the null hypothesis of the goodness-of-fit test is not rejected at the 1 % level. This applies both to the set of settlements mentioned before 1350 AD and all Medieval settlements. The least-cost distance of the Medieval settlements to the streams surrounded by fertile valleys is significantly lower than that of the reference distribution (Fig. 3, right). For $\alpha = 1\,\%$ the null hypothesis is rejected both for the settlements mentioned before 1350 AD and all Medieval settlements. This test supports the model that Medieval settlements primarily developed in the fertile valleys rather than close to ancient trade routes.

5 Discussion

The Kolmogorov–Smirnov test relies on the assumption that the sample is iid. There are some indications that Medieval settlement locations are not quite iid: (a) the distance between any two neighbouring settlements is above a certain limit (no settlement is on top of another one); (b) Siegmund (2009) discusses early Medieval settlements in the fairly flat area of the Niederrhein, Germany, and comes to the conclusion that they are regularly spaced with a distance between 2.25 and 3.25 km. However, for $n > 40$ the Kolmogorov–Smirnov test is conservative tending

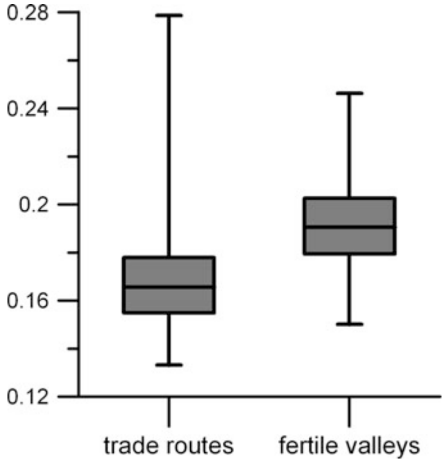

Fig. 4 Box plots of the autocorrelation coefficients (Moran's I) of 190 sets of random points in the study area

to favour the null hypothesis, and a very low confidence level was chosen. Therefore the conclusion that fertile valleys were preferred probably is still valid.

To check the effect of autocorrelation, Fotheringham et al. (2000) suggest comparing the autocorrelation coefficient of the data set investigated with those of experimental distributions. Their approach assumes that the attribute values are known only at the given spatial locations. But at any point within the study area the attribute value can be derived from the accessibility map, so a different approach is applied: 190 data sets each consisting of 513 random points within the study area were simulated and the autocorrelation coefficients (Moran's I) for the distance to trade route (or fertile valleys respectively) were calculated for each data set. The popular Moran's I measures the spatial dependency of nearby locations (Fotheringham et al. 2000):

$$I = \left(\frac{n}{\sum_i \sum_j w_{ij}} \right) \left(\frac{\sum_i \sum_j w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_i (x_i - \overline{x})^2} \right) \quad (4)$$

where n is the number of samples, x_i is the value at location number i, and the weights w_{ij} are determined by the inverse distance between two locations i and j. As calculating least-cost distances is very time-consuming, the Euclidian distance was used instead. If neighbouring values are more similar than with a random distribution, a positive autocorrelation coefficient is expected. Negative autocorrelation indicates higher dissimilarities of nearby locations than with a random distribution.

According to the box plots of the Moran's I experimental distributions (Fig. 4), for Medieval settlements both Moran's I for trade routes (0.15079) and fertile valleys (0.16871) can be found in the lowest quartile of the distributions. In fact, for fertile valleys, the coefficient is within the bottom 10 % range, and for trade routes within the bottom 20 % range. This is probably the result of the fact that the settlements are more regularly distributed than a random sample. A more positive autocorrelation than expected should create a problem with the test performed above, however, with the observed fairly low values of Moran's I, no negative impact on the test results is expected.

After performing the tests and on checking the digitized routes against the description by Nicke (2001), it was found that a route (Homburgische Eisenstraße) was omitted. So on including this route, the ancient route model might become significant. However, according to Nicke (1995) only two ancient routes within the study area played an important role for the spread of Medieval settlements (Brüderstraße and Zeitstraße). The Zeitstraße seems to be very problematic, because it is difficult to reconstruct by LCPs and because there are only few settlements in its vicinity. So maybe any other set of ancient trade routes was more important with respect to the settlement history. Testing all subsets of trade routes or each trade route individually may help to clarify if any of these old roads played a significant role with respect to Medieval settlement location in the study area.

References

Conover, W. J. (1971). *Practical Nonparametric Statitistics* (pp. 293–301). New York: Wiley.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative Geography. Perspectives on Spatial Data Analysis* (pp. 201–206). London: Sage.

Goodchild, M. F., & Hunter, G. J. (1997). A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science, 11*(3), 299–306.

Grewe, K. (2004). Alle Wege führen nach Rom - Römerstraßen im Rheinland und anderswo. In H. Koschik (Ed.), 'Alle Wege führen nach Rom', Internationales Römerstraßenkolloquium Bonn. *Materialien zur Bodendenkmalpflege im Rheinland* (Vol. 16, pp. 9–42).

Herzog, I. (2010). Theory and practice of cost functions. In F. J. Melero, P. Cano, & J. Revelles (Eds.), *Fusion of Cultures. Abstracts of the XXXVIII Conference on Computer Applications and Quantitative Methods in Archaeology, CAA 2010* (pp. 431–434), Granada.

Kolodziej, H. (2005). *Herzog Wilhelm I. von Berg. 1380–1408*. Bergische Forschungen 29. VDS – Verlagsdruckerei Schmidt GmbH, Neustadt an der Aisch, 12.

Llobera, M., & Sluckin, T. J. (2007). Zigzagging: Theoretical insights on climbing strategies. *Journal of Theoretical Biology, 249*, 206–217.

Minetti, A. E. (1995). Optimum gradient of mountain paths. *Journal of Applied Physiology, 79*, 1698–1703.

Nicke, H. (1995). *Das Oberbergische Land. Ein Landschaftsportrait*. Wiehl: Martina Galunder Verlag.

Nicke, H. (2001). *Vergessene Wege*. Nümbrecht: Martina Galunder Verlag.

Pampus, K. (1998). Urkundliche Erstnennungen oberbergischer Orte. *Beiträge zur Oberbergischen Geschichte. Sonderband*, Gummersbach.

Siegmund, F. (2009). Archäologische Beiträge zur Geschichte des frühen Mittelalters im Raum zwischen Niers und Kendel. Weeze und der Raum an Niers und Kendel im Mittelalter. *Weezer Archiv, Schriftenreihe der Gemeinde Weeze 3*, 136–144.

Tobler, W. (1993). *Non-isotropic geographic modeling*. Technical Report 93–1. http://www.geodyssey.com/papers/tobler93.html (seen 06-September-2012).