TAMÁS DUSEK

Bidimensional Regression in Spatial Analysis*

Bidimensional regression is a method developed by Waldo Tobler for comparing the degree of resemblance between two two-dimensional configurations of points or surfaces. In case of spatial analysis and any other research focusing on two- or multidimensional configurations, a need arises to compare these with each other. Bidimensional regression makes it possible in such a way that it transforms one of the configurations of points being in different coordinate systems into the coordinate system of the other with the proper degree of displacement, rotation or rescaling. Between the points of configurations transformed into a joint coordinate system this way, it is possible to determine the degree of the local and global similarity or dissimilarity of the configurations.

After the antecedents in the sixties and seventies, Waldo Tobler published his study introducing this method in 1994[1] (Tobler 1961, 1965, 1978, 1994). In his demonstrative example, Tobler compared a medieval map[2] of Britannia with a modern map. He mentioned further application options, for example comparisons between the faces of father and son, shapes of leaves, skill forms of the Australopithecus and the Homo sapiens as well as signatures. The possible applications of this technique cover such a large range that the procedures should become as well known as the technique of ordinary regression. (Tobler 1994, 187) Lloyd and Lilley also used this method to analyse the Gough map.

In case of point configurations generated with a multidimensional scaling, if they represent a matrix containing non-air kilometre distances between different geographical points, as well as at cognitive maps, a need arises to compare it with geographical coordinates. Because of the differences (displacement, rotation and scaling) between the geographical and the multidimensional scaling (hereinafter: MDS) coordinate systems, the configurations cannot be directly compared, however, bidimensional regression makes it possible just by the objective fitting of the two coordinate systems with the smallest possible defect. In connection with the comparability of time and geographical spaces, Ewing wrote in 1974 that there was no such method in existence (Ewing 1974, 165); however, the study of Ahmed and Miller of 2007 had already used this method.

This paper presents the main characteristics of the bidimensional regression mostly by the help of a demonstrative example. Since unidimensional, bivariate regression is a well

* This study was made with support of the Bolyai János Research scholarship. The study is the edited version of the presentation delivered at the session of “Analytical methods with space parameters” of the HAS RSC Research Methodology Subcommittee on 28 September 2010.

1 In addition to the study of Waldo Tobler of 1994, the study of Friedman and Kohler of 2003 focuses on the general issues of this method. In the demonstrative example of this latter study, a cognitive map is compared with a topographic map. However, there is no description of this method in the books of spatial statistics or general statistics.

2 Gough map.

known and widely used analytical method, I will dwell on the resemblances and differences between the bidimensional regression and the unidimensional regression.

Signs and abbreviations

In a unidimensional regression, the closeness of the relationship between different characteristics belonging to the same observational units is under investigation, while in case of a bidimensional regression, the degree of the relationship between two configurations that can be mutually associated with each other. In a unidimensional regression, the possibility to associate data is obviously ensured by the fact that they belong to the same observational units. In case of a bidimensional one, in geographical applications, the observational units are the same, but the distances between them may be different (e.g. air kilometre distance and time distance). In non-geographical applications, some kind of procedure is applied to make the points of two point patterns with the same number of points, e.g. the coordinates of specific points of two leaves, mutually connectible. The configurations may represent real point configurations and plains represented by points.

The point configurations to be compared may be displayed in two or one coordinate system(s) (figure 1.a and 1.b). The b part of the figure is confusing to some extent as the starting-point, the scaling and the direction of the coordinate systems may be arbitrary. To be more explicit, one part of the task is to ensure the possibility to display configurations of different coordinate systems in a single coordinate system. If X_i, Y_i are the coordinates of the independent configuration, A_i, B_i the coordinates of the points of the dependent configuration, then A_iB_i will be the coordinates of the independent configuration displayed in the coordinate system of the dependent configuration (Table 1). If we would like to express MDS coordinates in geographical coordinates, then the MDS configuration must be chosen as an independent configuration and the geographical configuration as a dependent one. In the knowledge of the relationship between the two configurations, on the one hand, we would like to determine the average degree of displacement, rotation and scaling; on the other hand, it is desirable that we would be able to give the transformed coordinates of those points where there was no observation.

Figure 1.a

The point configurations to be compared are consisted of points that can be mutually associated

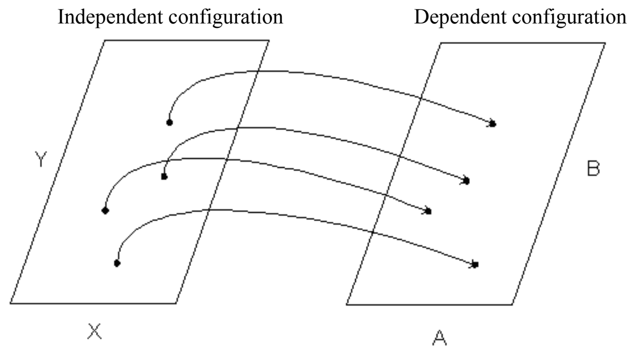
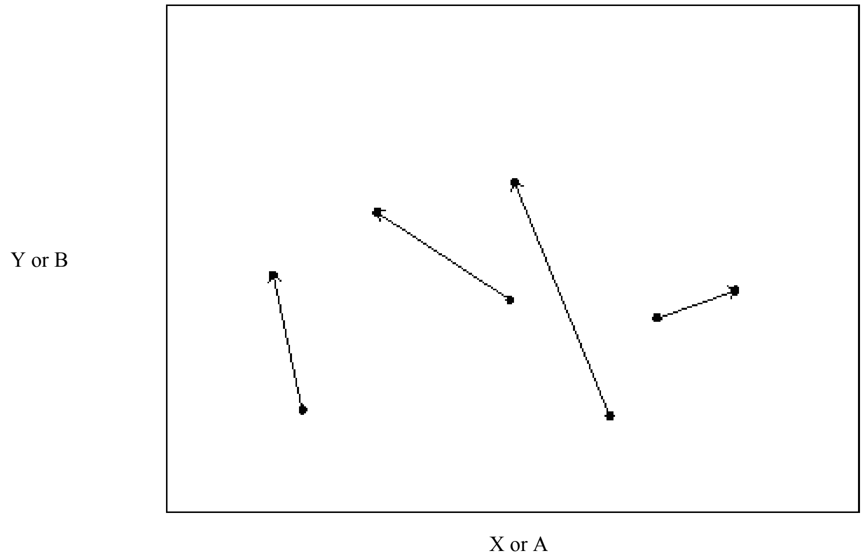


Figure 1.b

Joint coordinate system showing the point configurations to be compared



Source: Own chart based on Tobler (1994).

Table 1

Marking the coordinates of two point configurations

XY (Independent configuration)	AB (dependent configuration)	$\hat{A}\hat{B}$ (independent configuration in the coordinate system of the dependent configuration)
X_1 Y_1	A_1 B_1	$\hat{A}_1\hat{B}_1$
X_2 Y_2	A_2 B_2	$\hat{A}_2\hat{B}_2$
X_3 Y_3	A_3 B_3	$\hat{A}_3\hat{B}_3$
.		
.		
.		
X_n Y_n	A_n B_n	$\hat{A}_n\hat{B}_n$

In a unidimensional regression, we would like to minimize the $\sum(y_i - \hat{y}_i)^2$ amount, while in a bidimensional regression the $\sum(X_i Y_i - \hat{A}_i \hat{B}_i)^2$ amount. Based on the method of transforming coordinates, Waldo Tobler mentioned four possible solutions: Euclidian, affine, projective and curvilinear transformation; of these, the first three are linear functions, i.e. what is a straight line in one of the coordinate systems will be a straight line in the transformed coordinate system as well. In the Euclidian version, x and y coordinates change to the same degree (proportion), in the affine one the x and y coordinates may change in different proportions, while in the projective one the scale, the shape, and the rotation may change from point to point. This study is restricted only to how the Euclidian version is determined. The calculation of the three other ones would be more complicated, while in the projective and the curvilinear ones, the interpretation of the results would be more difficult as well.

Computation and interpretation of the parameters of the Euclidian regression

Equations relating to the computation of the Euclidian version can be seen in Table 2. Of the markings, x and y are the coordinates of the independent configuration, a and b the coordinates of the dependent configuration, while $\hat{a}$ and $\hat{b}$ the coordinates of the independent configuration inside the coordinate system of dependent configuration. Of the four parameters of the first (matrix-algebraic) equation, α1 is to determine the degree of the horizontal displacement, while α2 the degree of the vertical displacement. If we draw a parallel between the one and the two dimensional regressions, then these two coefficients will correspond to the β1 (constant) parameter of the one dimensional regression. β1 and β2 are to determine the scale difference (Φ) and the angle of the rotation (Θ) in a way that can be seen in the first and second equalities. If Φ=1, then there is no scale difference between the two configurations, if Φ>1 then it means the zooming of XY and the reduction of that at Φ<1. If Θ=0, then there is no need to rotate the XY coordinate system, if it is negative, then it means a clockwise rotation. Since the arcus tangent equation can be interpreted only between -90 degrees and +90 degrees, 180 degrees must be added to Θ, if $\beta_1<0$. Φ is the β_1 parameter of the one-dimensional regression, Θ is the specific parameter of the two dimensional case.

Table 2

Equations of a bidimensional Euclidian regression

1. Equation of the regression	$\begin{pmatrix}\hat{A}\\ \hat{B}\end{pmatrix}=\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}+\begin{pmatrix}\beta_1 & -\beta_2\\ \beta_2 & \beta_1\end{pmatrix}*\begin{pmatrix}X\\ Y\end{pmatrix}$
2. Scale difference	$\Phi=\sqrt{\beta_1^2+\beta_2^2}$
3. Rotation	$\Theta=\tan^{-1}\left(\frac{\beta_2}{\beta_1}\right)$
4. Calculation of β_1	$\beta_1=\frac{\sum(a_i-\bar{a})*(x_i-\bar{x})+\sum(b_i-\bar{b})*(y_i-\bar{y})}{\sum(x_i-\bar{x})^2+\sum(y_i-\bar{y})^2}$
5. Calculation of β_2	$\beta_2=\frac{\sum(b_i-\bar{b})*(x_i-\bar{x})-\sum(a_i-\bar{a})*(y_i-\bar{y})}{\sum(x_i-\bar{x})^2+\sum(y-\bar{y})^2}$
6. Horizontal displacement	$\alpha_1=\bar{a}-\beta_1*\bar{x}+\beta_2*\bar{y}$
7. Vertical displacement	$\alpha_2=\bar{b}-\beta_2*\bar{x}-\beta_1*\bar{y}$
8. Correlation based on random errors	$r=\sqrt{1-\frac{\sum\left[(a_i-\hat{a}_i)^2+(b_i-\hat{b}_i)^2\right]}{\sum\left[(a_i-\bar{a})^2+(b_i-\bar{b})^2\right]}}$
9. Breaking down the sum of squares	$\sum\left[(a_i-\bar{a})^2+(b_i-\bar{b})^2\right]=\sum\left[(\hat{a}_i-\bar{a})^2+(\hat{b}_i-\bar{b})^2\right]+\sum\left[(a_i-\hat{a}_i)^2+(b_i-\hat{b}_i)^2\right]$ SST=SSR+SSE
10. Calculation of $\hat{a}$	$\hat{a}=\alpha_1+\beta_1(X)-\beta_2(Y)$
11. Calculation of $\hat{b}$	$\hat{b}=\alpha_2+\beta_2(X)+\beta_1(Y)$

Source: Based on Tobler (1994) and Friedman–Kohler (2003). A key to these signs and abbreviations can be found in the text.

The greater the similarity between the two point configurations, the greater the degree of the bidimensional correlation (r). If as a result of displacement, rotation, and scaling, the coordinates of the points will correspond to each other, the indicator will reach the maximum value at one. The minimum value of the correlation is zero, which means that all points of one of the point configurations have the same coordinates. At this time, the centre of gravity of the two configurations will be the same, but the distance between them will be the same as the distance from the centre of gravity of that configuration, which is not clustering around one point. It may happen that we would like to disregard one of the three transformations, in that way the calculations relating to the similarity between the two configurations might also be implemented.

In principle, the breaking down of the sum of squares is performed in the same way as in case of the unidimensional one, the abbreviations are also the same (SST: total sum of squares, SSR: the sum of squares explained by the regression, SSE: residual sum of squares, which is not explained by the regression). The practical and interpretational difference is indicated by the fact that in a bidimensional regression, the difference indicates the distance from the centre of gravity of the analysed points and not from the average of a quantitative attributum variable.

In tables 3 and 4 as well as in figure 2, an example can be seen showing a comparison between the geographical distances and the railway kilometre distances of five settlements. Railway kilometre distances determine a non-Euclidian space, a two dimensional scaling was applied to their bidimensional approximation. (For example, it shows that as Mosonmagyaróvár may be reached by only a roundabout route first of all from Sopron and, to a smaller extent, from Szombathely and Pápa, therefore it will be farther away from the previously mentioned three settlements in the space of the railway network). The average of the MDS coordinates is zero, they have no unit of measurement, and the geographical coordinates are the coordinates of the uniform national projection given in terms of kilometres. The value of Φ of 62.41 gives the exact difference of their scale. MDS coordinates must be rotated by 19.347 degrees in an anticlockwise direction (value of Θ). The degree of the horizontal shift is 504.59 (α_1), and 247.96 is that of the vertical one (α_2).

Table 3

Two-dimensional regression between the space of the railway network and the geographical space

Settlement	MDS-coordinates		Geographical coordinates		MDS-coordinates	
	X	Y	A	B	$\hat{A}$	$\hat{B}$
Győr	0.350	0.055	543.90	260.70	524.10	258.40
Mosonmagyaróvár	0.578	0.354	516.60	280.00	531.30	280.70
Pápa	0.273	–0.540	530.00	223.10	531.80	222.00
Sopron	–0.600	0.456	465.30	262.20	459.80	262.4.0
Szombathely	–0.601	–0.330	467.10	213.80	476.00	216.20
Average	*0.000*	*0.000*	*504.58*	*247.96*	*504.60*	*247.94*

Table 4

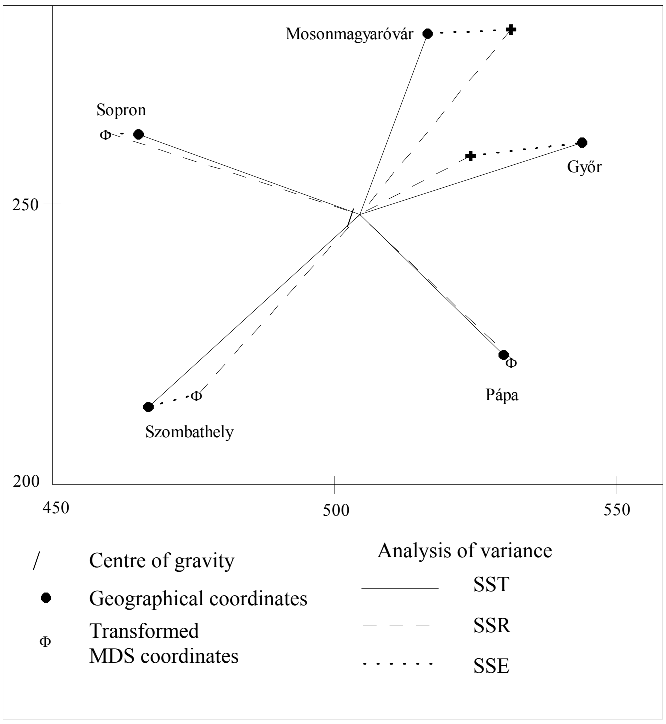
Results of the two dimensional regression (concerning data of table 3)

r	r^2	α_1	α_2	β_1	β_2	Φ	Θ	SST	SSR	SSE
0.956	0.913	504.59	247.96	58.877	20.673	62.41	19.347	8462.9	7730	732.9

The utility of the correlation and the determination coefficients is the same as in the case of the unidimensional regression, however, as it is well known in this latter case, it should be cautiously interpreted since its value may be influenced by many kinds of factors. The comparison of random errors of observations shows those points, which are most responsible for the deviations, though the warning that the influential observations are not necessarily the most outstanding ones is also valid in this case. The shift in the direction of the centre of gravity means that the average distance of the given point from the other points is smaller in the independent configuration, than in the dependent configuration. For example, in figure 2, the average distance between Győr and Szombathely, regarding its proportions, is smaller, while that of the other three settlements is greater in the space of the railway network, than based on air kilometre distances. (The railway network distances, in an absolute way, are also greater in case of Győr and Szombathely than the air kilometre distances, but they increased at a lower rate compared with the three other settlements.)

Figure 2

Geographical distances of settlements and their transformed MDS coordinates representing railway network distances



The independent point configuration with its original coordinates cannot be seen in figure 2, as the great differences of the scale and the displacement would not make possible the joint representation of XY and AB. The relative configuration of XY and AB and the proportions between the distances of certain points correspond to each other. The results gained with the transposition of dependent and independent configurations can be seen in table 5. The closeness of the correlation do not change, the absolute values of the transformations will be different; instead of extension, a reduction corresponding to the degree of the reciprocal value of the extension, shifts of opposing direction and of different degree and a rotation of opposing direction (but of the same degree) will be necessary.

Table 5

Results of a bidimensional regression
(to the data of table 3, but XY and AB are transposed)

r	r^2	α_1	α_2	β_1	β_2	Φ	Θ	SST	SSR	SSE
0,956	0,913	–8,171	–0,978	0,014	–0,00485	0,01464	–19,347	1,9851	1,8132	0,1719

Comparison with other procedures

Several other methods exist to compare two dimensional configurations: various shape indices, separate regressions by dimension, correlation between the distance matrices of points. The advantage of the two-dimensional regression as opposed to all existing methods, is that this is the only method that takes into account all information relating to the configuration of the formations. The separate regression by dimension is a unidimensional method, it does not take into account that XY and AB express the coordinates of certain points in an inseparable way. Regressions by dimension are also sensitive to rotation (or the lack of that). The regression between the distance matrices, though it provides another indicator for the degree of similarity, is not appropriate to determine scale difference, rotation and displacement.

The coordinates of the configurations to be compared may derive from a bidimensional scaling. However, the different local and global indicators (the Stress measures) of the goodness of the two-dimensional scaling do not reveal a deviation in relation to the geographical space, but the differences of the original distance matrix and the two-dimensional distance matrix. Therefore, these two methods may be connected inside the same analysis, but in a mutually supplementing and not substituting manner.

Toolkit of visual presentation and an example on the application of this method

This method is not included in any statistical or geostatistical software. However, the coefficients themselves can be easily computed in Excel based on the formulas. Mapping is much more labour intensive. The software, originally developed by Waldo Tobler, then redeveloped by Guerin, which can be downloaded from the following website: http://www.spatial-modelling.info/Darcy-2-module-decom-paraison is a help in this. The coefficients also can be calculated with this; however, the applicability of the visual display is significantly reduced by the fact that the figures are non-editable and non-inscribable.

I will show through examples the tools of the visual display relating to this method including some further minor questions. Concerning our 23 towns (Budapest and towns of county rank with the exception of Érd), table 6 is to show the results of the relationships between coordinates received with bidimensional scaling using geographical coordinates and corresponding distance matrices determined in different ways. Of the options offered by the Elvira internet time table search engine, I took the shortest travel time as a railway time distance, the public road distance matrices were made and put at my disposal by Péter Tóth, who used a piece of Microsoft MapPoint 2009 software (Széchenyi István University). Public road data are relating to July 2008, the railway data to November 2009.

Table 6

Results of bidimensional regressions between geographical coordinates and coordinates determined by multidimensional scaling from other different distance matrices (22 towns of county rank and Budapest)

Denomination	r	r^2	α_1	α_2	β_1	β_2	Φ	Θ
Railway network distance	0.974	0.949	649.0	199.5	188.3	40.2	192.6	12.06
Railway network time distance	0.941	0.885	649.0	199.5	184.9	28.3	187.0	8.71
Railway cost distance	0.973	0.947	649.0	199.5	191.0	29.9	193.3	8.91
Public road distance	0.990	0.980	649.0	199.5	192.8	32.7	195.5	9.63
Public road time distance	0.975	0.950	649.0	199.5	–190.6	–28.7	192.8	188.57

Remark: Railway distances: November 2009, public road distances: July 2008.

The relatively high level of correlations indicates that at this level there is no such significant distortion in the railway, public road network and time space, which may significantly alter the neighbourhood relationships of the settlements. If we analyse the more detailed network of 142 settlements, then the picture is modified at the time distances (Table 7). In the selection of the 142 settlements, the settlement size and the railway network location played a role. In those configurations, which are made up by many points, because of the crowdedness of the chart, it is more difficult to see the shifts, but in case of a good fitting, it is easy to interpret the overall view. However, many points and bad fitting jointly result in such charts, which can be interpreted in a more difficult way, though the more interesting single shifts may be stressed in this case as well and the spatial segment especially responsible for the bad fitting might also be identified.

Table 7

Results of two-dimensional regressions between geographical coordinates and coordinates determined by multidimensional scaling from other different distance matrices (142 Hungarian settlements)

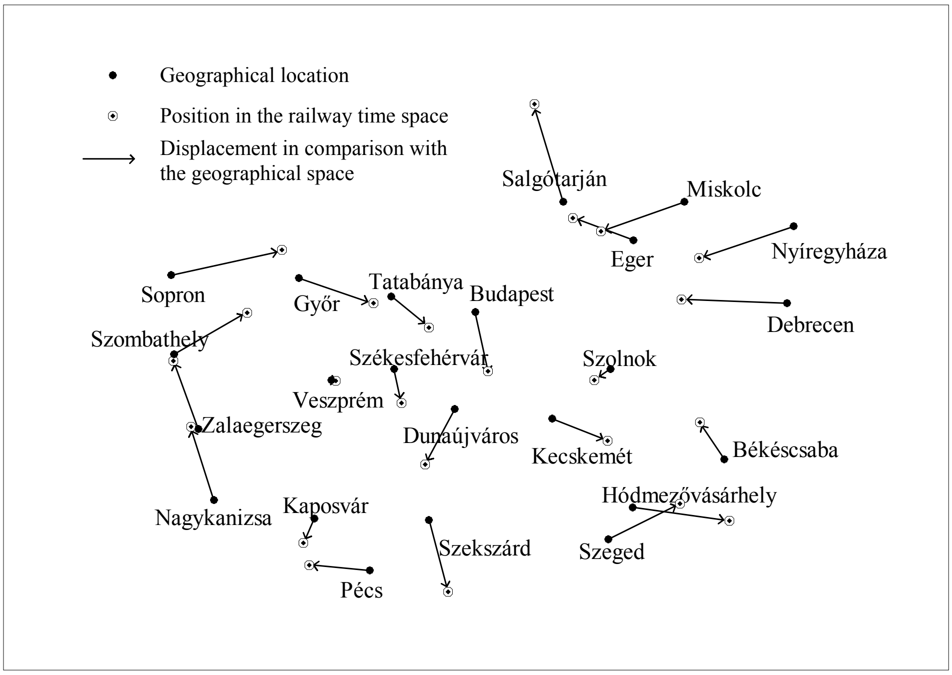
Denomination	r	r^2	α_1	α_2	β_1	β_2	Φ	Θ
Railway network distance	0.984	0.968	687.0	216.0	–187.4	–60.5	197.0	197.9
Railway time distance	0.781	0.610	687.0	215.9	–155.3	–33.7	158.9	192.2
Public road time distance	0.446	0.199	687.0	216.0	–84.4	–30.1	89.6	199.6

Remark: Railway distances: November 2009, public road distances: July 2008.

The geographical and railway time spatial position of the examined settlements can be seen in Figure 3. Graphical display is much richer in information than the quantitative display of the coordinates and their differences, because it shows the size and the direction of the change concerning all settlements. For example, it can be seen that Budapest shifted in the direction of the centre of gravity, because its railway accessibility is better than its otherwise favourable, near-central geographical location. Nyíregyháza, Debrecen, Miskolc, Győr, Sopron, Tatabánya significantly shifted in the direction of the centre of gravity, and, to a smaller extent, the same can be said of Szombathely and Békéscsaba. The settlements of Southern Great Plain and Southern Transdanubia moved farther away from each other, because of the non-appropriate east-west interconnection of the south of the country. The time spatial position of Dunaújváros and Salgótarján is significantly worse than their geographical positions. Gábor Szalkai draw the same conclusions through the use of detour indices concerning the differences of the domestic railway time space and the geographical space (Szalkai 2001, Szalkai 2004), however, the method of the detour index is not appropriate in itself to indicate those differences that can be determined from direction vectors.

Figure 3

Location of towns of county rank and Budapest based on geographical coordinates and railway time spatial coordinates



Remark: Figure 3–5 were made with MapInfo based on computed, modified coordinates.

At the same time, the figure contains only those points that take part in the regression, so it is more difficult to read the information, similarly to that as if only the estimated points would be displayed in a unidimensional two-variable regression instead of the straight line of the regression. Interpolation makes it possible to display those points that do not take part in the regression if we know the shift of those points that take part in the regression. The mesh fitted on the coordinate system of the dependent configuration and the interpolated, modified position thereof further generalize the information derived from those points that take part in the regression. In parts a and b of figure 4, in addition to the original and the interpolated mesh, those points that take part in the interpolation (towns of county rank) can also be seen. However, it is very rare that there are no linear geographical objects that are well known from the topographical maps, for example in case of Hungary county borders, the position of which may be interpolated as well (figure 5). This figure spectacularly displays the direction and degree of the more significant differences between the geographical space and the railway time space.

Figure 4.a

Geographical coordinates of a grid

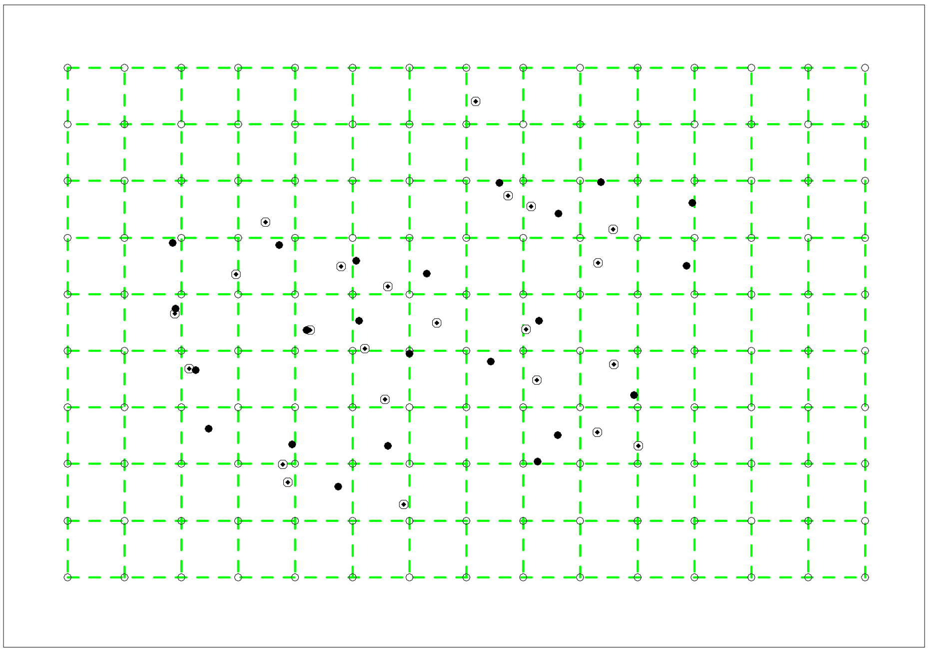


Figure 4.b

Warped grid: time space coordinates of the grid

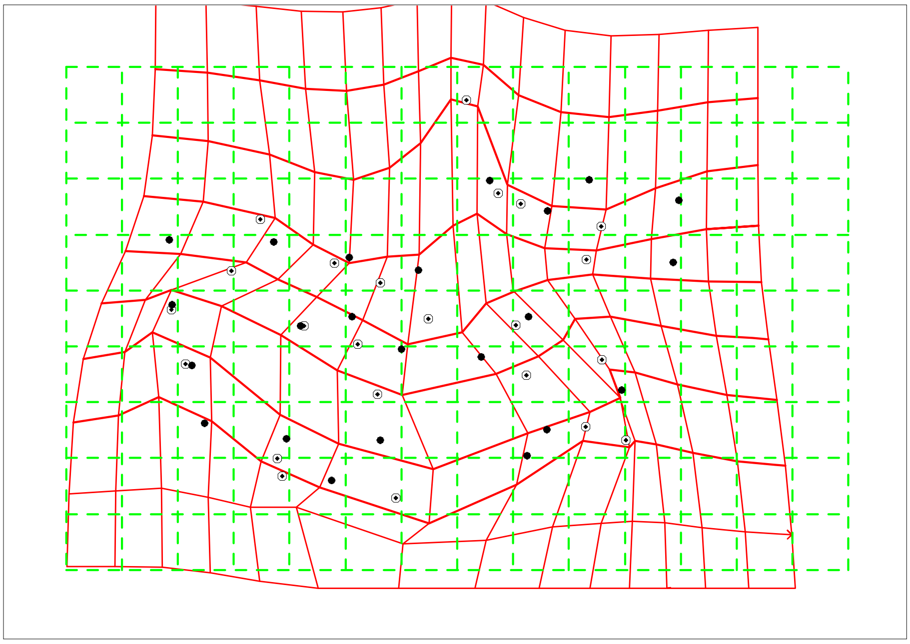


Figure 5

Interpolated position of county borders in railway time space

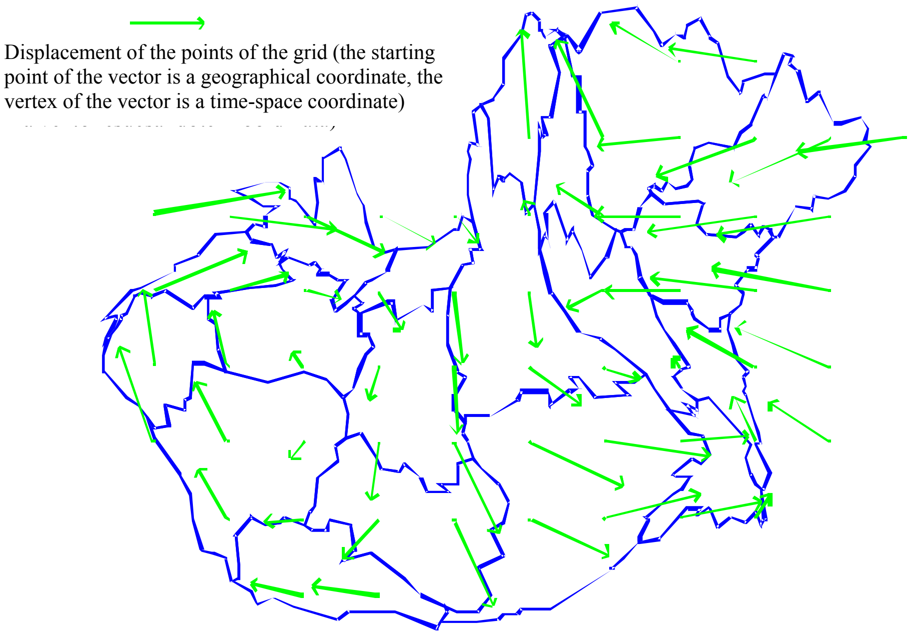
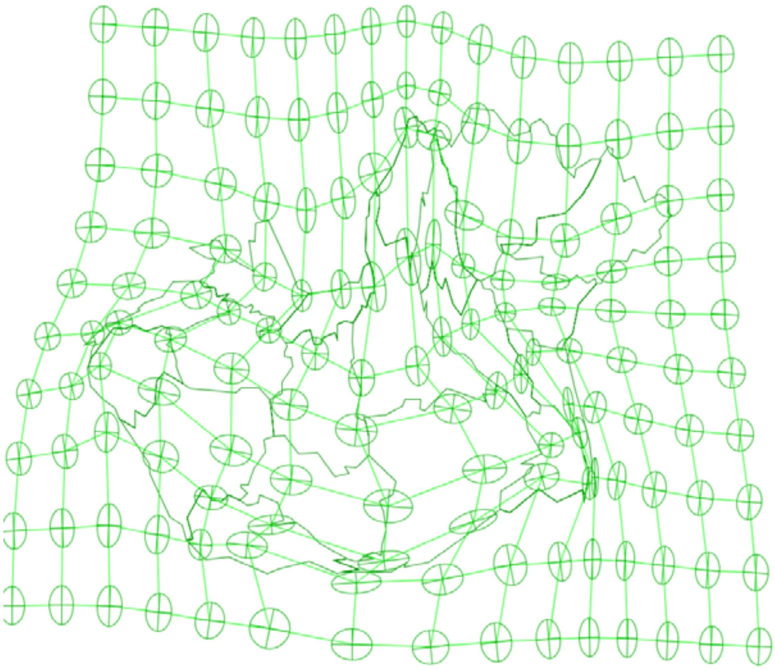


Figure 6

Distortion of the railway time space in comparison with the geographical space



Remark: This figure was made with the Darcy software (by B. Guerin). This software can be downloaded from the following website: http://www.spatial-modelling.info/Darcy-2-module-de-comparaison.

In the cartography, the distortion of the geographical projection may be expressed in a graphic manner with the help of the Tissot's indicatrix. Waldo Tobler presented, as an analogy of this, that the degree of the local deviation of the bidimensionally-compared configurations can be displayed with distortion ellipses (Tobler 1994). These distortion ellipses can be seen in the nodes of the interpolated grid on figure 6. If the area of the ellipse is greater than average, then in the given spatial segment, the time space will be more elongated than the geographical space. The main axis of the ellipse will shift in the direction of the greatest distortion.

Concluding thoughts

As it can be seen from the presented calculations and display options, bidimensional regression provides basic data for the visualization, interpretation, measurement, and analysis of the non-Euclidian spaces, and it is not only a procedure of comparing configurations with each other, which is much more effective and rich in information compared with the other known methods. If the coordinates are already available (the determination of these can be very laborious if the distance data needs to be collected on an individual basis), the calculation of correlation and regression is not complicated; in the visual display of the results only the display of the interpolated coordinates is more labour intensive. Then again, in the relation to this method, a number of questions of detail might be discussed, among them it should be stressed, the development of indicators that can be used to characterize the influential observations, the connection with the multidimensional scaling and the presentation of the affine transformation as well as the interpolation procedures. The present study's intention is to draw attention to the existence of this method in the first place, with the aim to facilitate its penetration and practical application.

REFERENCES

Ahmed, N. – Miller, H. J. (2007): Time-space Transconfigurations of Geographic Space for Exploring, Analyzing and Visualizing Transportation Systems. *Journal of Transportation Geography* 15 (1): 2–17.

Ewing, G. (1974): Multidimensional Scaling and Time-space Maps. *Canadian Geographer* 18 (2): 161–167.

Friedman, A. – Kohler, B. (2003): Bidimensional Regression: Assessing the Configural Similarity and Accuracy of Cognitive Maps and Other Two-Dimensional Data Sets. *Psychological Methods 8(4):468–491.*

Lloyd, C. D. – Lilley, K. D. (2009): Cartographic Veracity in Medieval Mapping: Analyzing Geographical Variation in the Gough Map of Great Britain. *Annals of the Association of American Geographers* 99 (1): 27–48.

Szalkai Gábor (2001): Elérhetőségi vizsgálatok Magyarországon. *Falu–Város–Régió* 3 (10): 5–13.

Szalkai Gábor (2004): A közlekedéshálózat fejlesztésének hatása az elérhetőség változására. Magyar Földrajzi Konferencia CD-kiadványa

Tobler, W. (1961): Map Transconfigurations of Geographic Space. PhD dissertation, University of Washington, Seattle

Tobler, W. (1965): Computation of the Correspondence of Geographical Patterns. *Papers of the Regional Science Association* 15 (1): 131–139.

Tobler, W. (1978): Comparisons of Plane Forms. *Geographical Analysis* 10 (2): 154–162.

Tobler, W. (1994): Bidimensional Regression. *Geographical Analysis* 26 (3): 186–212.

Keywords: bidimensional regression, regional analytical methods, multidimensional scaling, railway time space.

Abstract

Bidimensional regression is a method developed by Waldo Tobler for comparing the degree of resemblance between two two-dimensional configurations of points or surfaces. It is an extension of linear regression where each variable is a pair of values representing a location in a two-dimensional space. Bidimensional regression numerically compares the similarity between two-dimensional surfaces through an index called bidimensional correlation. The aim of the study is the general description of the method, and to present examples of its application with the help of some real data, that of the Hungarian railway time space.