Optimal bandwidth selection in geographically weighted factor analysis for education monitoring problems

A.Timofeeva[1], K.Tesselkina[1]

[1]*Novosibirsk State Technical University, 20 Prospekt K. Marksa, 630073, Novosibirsk, Russia*

Abstract

Geographically weighted models are widely used for analyzing the spatial data. There is a problem with spatial data processing of extracting a potentially lower number of unobserved variables while a set of correlated variables is observed. The factor analysis is commonly used to overcome this problem. A bandwidth selection is a main difficulty during the identification a spatial heterogeneity of factor loadings. In the paper an original bandwidth selection criterion is proposed. It is based on the testing the difference between factor loadings of global and geographically weighted model. Using the simulated data it is shown that the criterion proposed allows to define accurately the appropriate number of nearest neighbors. The proposed approach is used to analyze real data on performance metrics of Russian universities.

Keywords: spatial data; geographically weighted factor analysis; bandwidth selection; factor loadings; nearest neighbors

1. Introduction

Large amounts of spatial data that need to be processed is stored in modern geographic information systems. Thus, the issue of reducing the dimension of the attribute space occurs frequently. The most popular approaches in this field are the method of the principal components analysis (PCA) and exploratory factor analysis (EFA).

Both the PCA and EFA, require homogeneity of observed data. This assumption is violated in cases where the observations depend on geographical factors. Thus, both similar and different degree of dependency between observed variables and latent factors in different geographic regions can be observed. Often some of its spatial properties are ignored in analyzing data process and standard methods for reducing dimension are used. However, such spatial effects are often important for a better understanding of investigated process, and PCA in this case may be replaced by geographically weighted PCA [1], when we want to explain a certain spatial heterogeneity in the data.

At the same time, the idea of building a geographically weighted model does not transferred so easy from the method of principal component to factor analysis. They have a number of substantial differences, and in fact, the purpose their use is different, in particular, factor loadings play the important role in the interpretation of the EFA results, reflecting the impact on the observed variables. For example, if there is a system of parameters, which are exposed to the same latent factor, the loadings on the main factor show the degree of impact on the indicators.

In this paper, we used the idea of building a local model of EFA for geographically nearest neighbors (adaptive bandwidth). However, there is the problem of choosing the number of nearest neighbors. Authors of paper [2] that is devoted to geographically weighted PCA, propose criterion ‘goodness of fit’, based on the minimization of the residual sum of squares. This makes sense, since the aim of principal component analysis is to present indicators in the space of smaller dimension with the least loss of information. Here, for the factor analysis, we suggest another criterion, which takes into account the significance of the differences of factor loadings of locally weighted and global models. This is more consistent with the aim of factor analysis.

Further, the principal component analysis and factor analysis are described in more detail and explained the difference between them. Essence of geographical weighting is presented, the problem of bandwidth selection is described. A new criterion for selecting the number of nearest neighbors is proposed. Advantages of this approach were demonstrated by comparison with the existing criterion of goodness of fit in the simulation study.

2. Geographically weighted variable reduction

PCA and EFA are both variable reduction techniques and sometimes erroneously considered as the same statistical method. However, there are distinct differences between PCA and EFA. Further, mathematical description of these approaches is given, and we explain how they are adapted to spatial data analysis.

2.1. Principal component analysis

There are n observations of m variables, so a data matrix X contains n rows and m columns. The columns in X are normalized with zero mean and unit variance. Then $R = X^{\mathrm{T}}X$ is the correlation matrix for X. The matrix X^{T} denotes the transpose of X. The matrix R is a real symmetric matrix and its factorization into a canonical form is

$$R = \Lambda\Phi\Lambda^{\mathrm{T}} \tag{1}$$

where an orthogonal matrix Λ contains the eigenvectors of R, and Φ is a diagonal matrix which entries are the eigenvalues of R. The eigenvalues of diagonal Φ imply the variances of the corresponding principal components. The eigenvectors in Λ are column vectors and represent the loadings of each variable on the corresponding principal component.

If the number of principal components equal to the number of variables, the decomposition (1) perfectly reproduces the correlation matrix R. By reduction m variables in q-dimensional sub-space ($q < m$) the correlation matrix is represented as

$$\hat{R}_q = \Lambda_q \Phi_q \Lambda_q^{\mathrm{T}}$$

where Λ_q denotes the matrix Λ with the first q columns, i.e. the loadings on the first q principal component, and Φ_q is a diagonal matrix which entries are the first q eigenvalues of R. The principal components are sorted in decreasing order of eigenvalues so the first principal components keep the most important information from the data set.

Component scores in q-dimensional sub-space are found by multiplying the original data matrix X by loading matrix Λ_q. The best rank q approximation to X is $\hat{X}_q = X \Lambda_q \Lambda_q^{\mathrm{T}}$. A standard result in linear algebra states that

$$\Lambda\Lambda^{\mathrm{T}} - \Lambda_q \Lambda_q^{\mathrm{T}} = \Lambda_{(-q)} \Lambda_{(-q)}^{\mathrm{T}}$$

where $\Lambda_{(-q)}$ denotes the matrix Λ with the first q columns removed.

To assess the quality of the reconstitution of X with q components, the dissimilarity between X and $\hat{X}_q$ is usually evaluated. The error matrix is

$$E = X - \hat{X}_q = X\Lambda\Lambda^{\mathrm{T}} - X\Lambda_q\Lambda_q^{\mathrm{T}} = X\Lambda_{(-q)}\Lambda_{(-q)}^{\mathrm{T}}.$$

The most popular coefficient used for evaluating the quality of PCA model is the residual sum of squares [3]

$$RESS_q = \left\| X\Lambda_{(-q)}\Lambda_{(-q)}^{\mathrm{T}} \right\| \tag{2}$$

where $\|M\|$ is the square root of the sum of all the squared elements of the matrix M.

Thus, mathematically, PCA depends on the eigen-decomposition of positive semidefinite matrices. Its main goal is to extract the important information from the data using the correlation between the variables and to represent it as a set of orthogonal principal components in the sub-space of lower dimension.

2.2. Factor analysis

EFA model assumes that the relationship between the measured variables is due to the effect of some unobservable (latent) factors. The input information is a correlation matrix R for all variables. It can be represented as [4]

$$R = \Lambda\Phi\Lambda^{\mathrm{T}} + \Psi \tag{3}$$

where Λ is a factor loading matrix reflecting the relationship between the variables and factors, Φ is a correlation matrix of q factors, Ψ is a covariance matrix of the unique factors.

The presence of unique factors in EFA model (3) is the main difference from the model of PCA (1). It is due to the fact that extracted latent factors do not fully (with some errors) describe the correlation between the observed variables. Uniqueness is the variance that is 'unique' to the variable and not shared with other variables. The independence of unique factors is assumed, so the matrix Ψ is a diagonal with uniqueness on the diagonal.

The matrices Λ, Φ, Ψ are estimated. In contrast to the PCA model the matrix Λ is of a particular interest, but loadings are not uniquely determined, so the rotation procedure is used, so that the resulting factor structure has a meaningful interpretation. With orthogonal rotation the independence between the latent factors is assumed. So matrix Φ is given as identity. There are a number of factor extraction methods for estimating loadings and uniqueness, for example, principal factor solution, minimum residual, maximum-likelihood method.

Minimum residual method is based on ordinary least squares (OLS). The loss function is

$$F_{OLS}(\Lambda, \Psi) = \mathrm{tr}\left(R - \left(\Lambda\Phi\Lambda^{\mathrm{T}} + \Psi\right)\right)^2. \tag{4}$$

Here $\mathrm{tr}(M)$ is the trace of a square matrix M. The OLS-estimates $\hat{\Lambda}$, $\hat{\Psi}$ are arguments at which the minimum of loss function (4) is achieved.

2.3. Geographically weighted models

The usage of local weighting as part of regression estimation initially was proposed by [5]. This approach is widely used in spatial data analysis [6] and known as geographically weighted model.

Geographically weighted model identifies spatial differences in the relationship between factors by constructing a regression model at each control point for geographically closed observations. The proximity regulated by assigning larger weights to closest points and reducing weights for observations as they move away from the control point. Thus, the weight is determined as a function of distance from the control point to the objects. The regression is estimated over the local subregion which volume is determined by the weight function parameter (a bandwidth).

Regardless of the form of weight function specified, the local correlation matrix is

$$R^{(i)} = \Lambda^{(i)} \Phi^{(i)} \Lambda^{(i)\mathrm{T}} \tag{5}$$

with respect to the local subregion of the i-th control point. The scores for the i-th control point on the m variables are $\mathbf{x}^{(i)}\Lambda^{(i)}$ where $\mathbf{x}^{(i)}$ is a vector of variable values at i-th control point.

Similarly geographically weighted EFA model is defined as

$$R^{(i)} = \Lambda^{(i)} \Phi^{(i)} \Lambda^{(i)\mathrm{T}} + \Psi^{(i)} \tag{6}$$

and values of matrices $\Lambda^{(i)}$, $\Psi^{(i)}$ are estimated with respect to the local subregion of the i-th control point.

3. Criteria of bandwidth selection

The choice of bandwidth value has a decisive influence on the estimation quality [6]. If someone takes bandwidth too large, then almost all observations will be included in the model, so it will be coincidental to the global model without geographical weighting. Thus it will not be possible to describe the change of the explanatory factors impact depending on the spatial location of the objects. Otherwise, too small bandwidth leads to the overfitting problem: the model perfectly predicts the training data, but drastically fails on some new datasets.

The choice of a bandwidth parameter cannot be based on the common fitting indicators (like R-squared, the mean square error, etc.). So for optimal bandwidth selection the cross-validation technique is often used when the training and quality evaluation both are produced on the distinct sample data [7]. There are some other approaches to solve this problem, for instance, Akaike information criterion [7], and the Lagrange multiplier test [8].

In recent studies on the bandwidth selection [8, 9] for geographically weighted models two essentially different methods to the weighting function construction are considered:

- with a fixed local area radius;
- with a given number of nearest neighbors.

The second one is considered to be adaptive because it allows adjusting to varying density of the spatial location of the objects. Thus in the neighborhood of one control point the objects may be more concentrated than for the other points where they are more distant from each other. In such cases the latching of the local area radius leads to the fact that for some control points the regression will be estimated on a very large number of observations, for the others – on very small.

We selected the adaptive approach. Therefore, the task is to determine the optimal number of nearest neighbors, taking into account features of EFA.

3.1. Goodness of fit statistic

The criterion ‘goodness of fit’ is proposed for geographically weighted PCA model construction (see [2]). The criterion is based on the minimization of the residual sum of squares. It is calculated by the formula (2) for a global model. For a local PCA model (5) the residual sum of squares at the i-th control point is

$$RESS_q^{(i)} = \left\| X^{(i)} \Lambda_{(-q)}^{(i)} \Lambda_{(-q)}^{(i)\mathrm{T}} \right\|$$

where a superscript (i) denotes the values that are calculated in a local subregion of the i-th control point. The values of the residual sum of squares are summed for all the control points to calculate the goodness of fit statistics:

$$GOF = \sum_{i=1}^{n} RESS_q^{(i)} .$$

A set of control points can be selected in different ways. Leave-One-Out Cross-Validation (LOOCV) is the simplest procedure for cross-validation. It suggests that only one observation is selected as a control point from the data set, while other observations are considered as a training sample. The procedure is repeated until all the objects will be alternately selected as control points. The advantage of this approach is a computational speed. Often the model structure based on LOOCV leads to the overfitting problem and the forecast error underestimation [10]. In our case the wrong selection of bandwidth parameter may cause such problems.

The more complicated procedure is a Monte-Carlo cross-validation (MCCV) [10]. It assumes that the whole sample is separated randomly into training and check samples. Nevertheless, this choice could increase the computational time. Therefore, we have chosen LOOCV procedure.

3.2. Test the difference between global and local factor loadings

There are some problems with usage the cross-validation technique for evaluating the quality of EFA models. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. The task of variable reduction is not quite a prediction. Of course, we can use the loss function $RESS_q$ for PCA, and $F_{OLS}(\Lambda,\Psi)$ for EFA. But goodness of fit is not so important for EFA, the loadings are more interesting. For this reason a new criterion for bandwidth selection is proposed. It is based on testing the difference between global and local factor loadings.

A statistical inference for comparing global and local factor loadings is based on the information about mean values of loadings and their standard deviation. We need replications of sample data to get this information. One way to get it is to take a sample of the same size n from the rows of data matrix X with replacement. Let we have L replications of sample data. For each l th replication we estimate loading matrix Λ_l of global EFA model (3) and loading matrix $\Lambda_l^{(i)}$ of the geographically weighted EFA model (6). We can calculate matrices containing the average values of loadings for all replications

$$\bar{\Lambda}_l = \frac{1}{L}\sum_{l=1}^{L}\Lambda_l, \bar{\Lambda}_l^{(i)} = \frac{1}{L}\sum_{l=1}^{L}\Lambda_l^{(i)}.$$

The k, j th elements of matrices $\bar{\Lambda}_l$ and $\bar{\Lambda}_l^{(i)}$ are denoted by $m\left(\lambda_{kj}\right)$ and $m\left(\lambda_{kj}^{(i)}\right)$.

Similarly, we can calculate the variance of loadings for all replications. Let us denote them as $v\left(\lambda_{kj}\right)$ and $v\left(\lambda_{kj}^{(i)}\right)$. So the test statistic comparing the means is well known. It is given by

$$SS_{kj}^{(i)} = \frac{m\left(\lambda_{kj}^{(i)}\right) - m\left(\lambda_{kj}\right)}{\sqrt{1/L}\sqrt{v\left(\lambda_{kj}^{(i)}\right) + v\left(\lambda_{kj}\right)}}. \quad (7)$$

We propose the significance test statistics for optimal bandwidth selection

$$SS = \frac{1}{n \cdot q \cdot m}\sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{m}\left|SS_{kj}^{(i)}\right|. \quad (8)$$

The maximum value of the significance test statistics (8) corresponds to the optimal number of nearest neighbors. On the one hand it is evident that for the global model (maximum number of nearest neighbors) the numerator of (7) will be minimal, and vice versa. So we would expect that the geographically weighted EFA model with the smallest number of nearest neighbors will be the best by test statistics (7). But on the other hand a small number of nearest neighbors results in very large loadings variation. Thus, the denominator of (7) will increase with a decrease of the number of nearest neighbors. Essentially the significance test statistics (8) is a trade-off between differences in the average factor loadings and their variation.

To calculate statistics (8) we need to compare multiple EFA models. These comparisons require columns of factor loading matrices to be properly aligned. However, the most rotation criteria do not uniquely define a factor loading matrix. This is referred to an alignment problem [4]. The most popular method for aligning a factor loading matrix against another is to minimize the sum of squared differences of factor loadings in the two matrices. Further, in simulation study, this approach was used. We compared the dissimilarity of loading column of global EFA model and one of local models, initial and with the opposite sign. Column reflection (an operation when the signs of values in column are replaced with the opposite) of original column was carried out in cases when reflected loading column corresponded to less value of sum of squared deviations.

There are some problems with the calculation of the statistics (8). Firstly, it is necessary to conduct the $L \cdot n$-fold factor analysis. With a large number of control points and repeated replications, this procedure takes a very long time. A smaller number of control points can be taken to reduce the calculation time. Another way is to replicate the sample using the jackknife method. However, the decrease in the number of replications may result in loss of the quality of an optimal bandwidth selection.

Secondly, factors can be extracted using various methods. There is a problem with the starting values during the optimization of the log likelihood. The uniqueness is technically constrained to lie in [0, 1], but there are some problems with near-zero values, and the optimization is typically done with a lower bound of 0.005. Sometimes it is unable to optimize the likelihood from certain starting value, because the algorithm does not converge. If we try to increase or decrease the lower bound for uniqueness during the optimization, it allows a solution to be converged. However, such lower bound selection is practically not efficient in the case of $L \cdot n$-fold factor analysis. So more simple factoring methods should be used.

The method of principal axes may be used in the cases when maximum likelihood solutions fail to converge. However, it is based on the iterative algorithm, so it works rather slowly. If the procedure of factor analysis is repeated many times, the speed

of implementation of the factoring procedure is very important. Therefore, optimization procedures are more preferable. In addition, they produce even better solutions for some examples. Minimum residual method based on OLS usually has no problems with convergence and tends to produce better solutions.

Further, two approaches to the bandwidth selection are compared in a simulation study. For identification of PCA and EFA models one can use statistical packages, for instance, the free software for statistical analysis R [11]. The function princomp {stats} performs a principal components analysis on the given numeric data matrix, function efa{EFAutilities} performs exploratory factor analysis. The algorithms of optimal bandwidth selection are implemented using R.

4. Results of simulation study

The main purpose of the simulation study is concluded in comparison of approaches mentioned above in precision of the bandwidth selection. A simple one-factor model was chosen. The factor F is standard normally distributed. It affects three variables x_1, x_2, x_3. So the EFA model has the form

$$\begin{cases} x_1 = b_1 F + \varepsilon_1, \\ x_2 = b_2 F + \varepsilon_2, \\ x_3 = b_3 F + \varepsilon_3 \end{cases} \quad (9)$$

where b_1, b_2, b_3 are factor loadings, $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are random errors. The simulated random error ε_i was chosen as a normally distributed variable with variance that equals to $0.8^2 - b_i^2$.

The case with certain local centers of object's concentration was considered for modeling spatial heterogeneity. The whole number of those centers was equal to six, and each of them was represented by a circle with the same number of observations. All center's locations were chosen randomly within the unit square $[0,1]^2$. The sample size was $n = 300$. The true value of number of nearest neighbors was 50. Different spatial location of objects towards the centers was set in two ways.

In Model 1, the radius of a circle with homogeneous observations was fixed. It was set to 0.05.

Model 2 has a distance from the objects to the center of the local area multiplied by a coefficient $1 + 0.2K$, where K is a serial number of the area, $K = 1, \ldots, 6$. The number of objects belonging to the *K*-th region was set to $29 + 6K$. This ensures that there are areas with varying density of spatial location of objects.

Equal loadings were set for all objects in the same local area. Their values are shown in Table 1.

Table 1. The true values of factor loadings.

Factor loadings	*K*=1	*K*=2	*K*=3	*K*=4	*K*=5	*K*=6
b_1	0.37	0.32	0.25	0.57	0.58	0.71
b_2	0.43	0.18	0.49	0.16	0.32	0.28
b_3	0.2	0.5	0.26	0.27	0.1	0.01

The total number of experiments was equal to 50 for each model. The values of both statistics *GOF* and *SS* were calculated on the simulated data for fixed number of nearest neighbors. The number of nearest neighbors was set from 20 to 150 with the step of 10. The optimal number of nearest neighbors based on *GOF* statistics was selected as argument of the goodness of fit statistics minimum constructed as a result of the LOOCV procedure. The optimal number of nearest neighbors based on *SS* statistics was selected as argument of the significance test statistics maximum constructed as a result of the MCCV procedure. The number L of replications of sample data was set to 100. The size of random sample with replacement was $n = 300$. The final results are presented in Table 2.

Table 2. The optimal number of nearest neighbors.

	Based on *GOF* statistics				Based on *SS* statistics			
	Q1	Median	Mean	Q3	Q1	Median	Mean	Q3
Model 1	42.5	75	84.2	137.5	30	40	50.6	67.5
Model 2	52.5	70	83.8	120	22.5	40	49.2	60

It is clearly seen that the significance test statistics proposed determines the number of nearest neighbors more accurately for both model 1 and model 2. The goodness of fit criterion gives an average value of optimal number of nearest neighbors of about 80, while the significance test statistics reaches a maximum value when the average number of nearest neighbors almost equal to

50. Thus, the use of the goodness of fit criterion leads to an explicit overestimation of the bandwidth. In addition, SS statistics also provide a more accurate determination of the number of nearest neighbors. The interquartile range of the optimal bandwidth for it is 30.5. While the optimal number of nearest neighbors, according to the GOF criterion, has an interquartile range of 95 and 67.5, that is 2-3 times more than the results of use the second criterion.

Figure 1 shows the resulting values of statistics for one of the samples. The true number of nearest neighbors for model 1 that equals 50 is indicated by a dashed line. For model 2 the true number of nearest neighbors varies according to local area from 35 to 65 with the step of 6. The average number of nearest neighbors is 50.

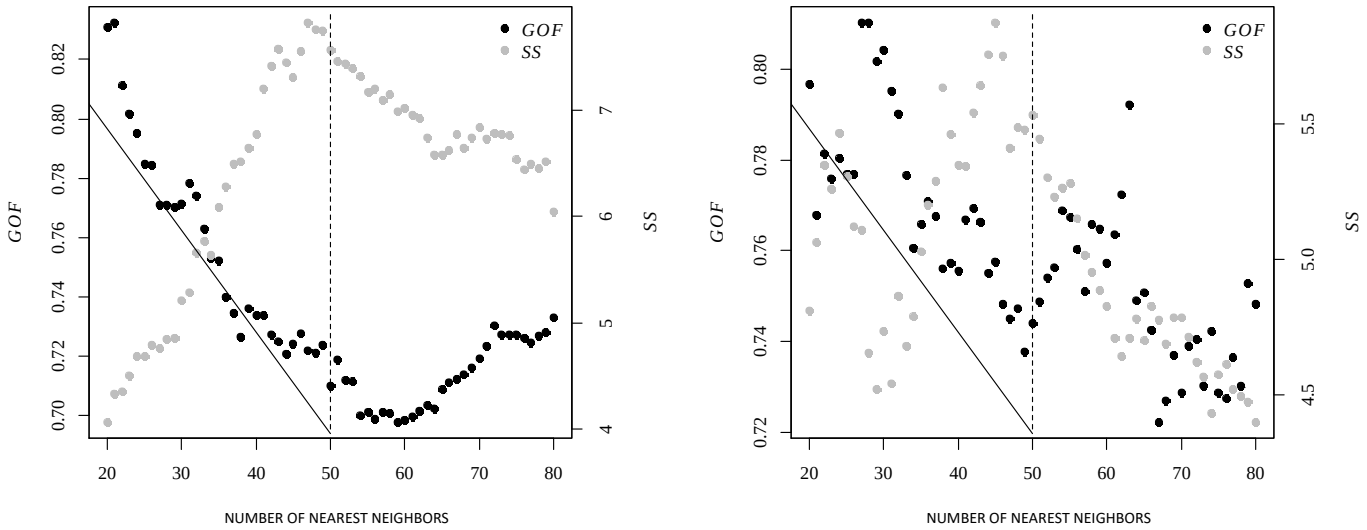

Fig. 1. The dependence on the goodness of fit statistics and the significance test statistics on the number of nearest neighbors for model 1 (a) and model 2 (b).

It should be noted that values of the goodness of fit statistics vary greatly. We see a lot of local minima and maxima. This fact complicates the use of optimization routines to find the best values of the bandwidth. At the same time, the dependence of the significance test statistics on the number of nearest neighbors appears smoother. This fact allows us to develop more effective optimization algorithms than direct-search method on the grid.

5. Application to educational monitoring

The Ministry of Education and Science of the Russian Federation initiated monitoring of the effectiveness of universities in 2012. Since then, all Russian universities are obliged to provide information on their activities on a number of indicators. The decision on the effectiveness of the university is made depending on whether the university is able to reach thresholds for most indicators. The leadership of different universities can differently determine the priority indicators. It is interesting to determine the structure of universities' efficiency taking into account regional differences in their activities.

We will rely on the model (9) for determining the structure of performance indicators. We are interested in the factor F that is the overall efficiency of the university. So the observed indicators of activities can be interpreted as

- x_1 is a financial and economic activity: income of the educational organization from all sources per one NDP;
- x_2 is a level of wages of the teaching staff: the ratio of the salary of PPP to the average wage for the economy of the region;
- x_3 is an employment of graduates: the proportion of graduates who have found employment during the calendar year following the year of release, in the total number of graduates of the educational organization who have studied the main educational programs of higher education.

Coefficients b_1, b_2, b_3 show the extent to which an particular performance indicator determines effectiveness.

The data from monitoring the effectiveness of educational institutions of higher education for 2015, downloaded from the pages of each individual university, were used as an information base [12]. 571 universities are represented in the sample. They provided information on performance indicators, branches of universities are not included in the analysis. The optimal number of nearest neighbors was chosen based on the SS statistics. The MCCV procedure was used. Control points were located in administrative centers within six federal districts: Central Federal District (CFD), Northwestern Federal District (NWFD), Volga Federal District (VFD), Ural federal district (UFD), Siberian Federal District (SFD), Far Eastern Federal District (FEFD). A total of 67 control points are set. The number L of replications of sample data was set to 300. The size of random sample with replacement was 571. The optimal number of nearest neighbors was 119. The average values of loadings for all replications denoted by m (b1), m (b2), m (b3) are presented graphically in Figure 2 using pie charts.

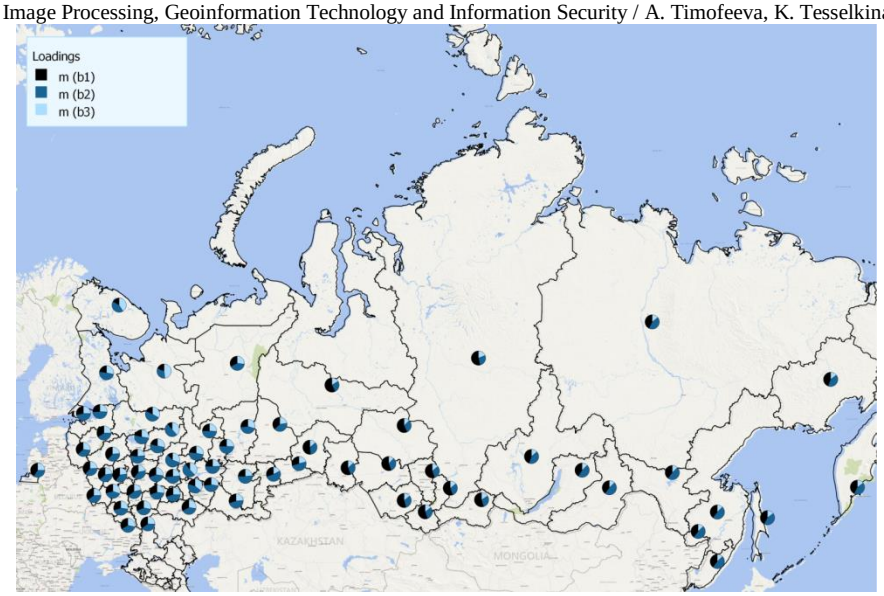

Fig. 2. Geographical variations of the factor loadings of universities' efficiency.

As can be seen from Fig. 2, the structure of indicators of the effectiveness of HEIs varies greatly depending on the territory. Thus, for the regions of Siberia and the Far East, the income of the educational organization has the greatest weight. For many European regions of the country, financial and economic activity does not have such a significant contribution. In most cases, the level of wages of scientific and pedagogical workers is most important in the formation of performance of universities. Only for one northern region of Russia (Arkhangelsk region), employment of graduates predominates in the structure of performance indicators Consequently we can conclude that for most universities the key performance indicators are financial, while the employment of graduates as a result of educational activities does not play such a significant role.

6. Conclusion

In the paper we propose an original criterion to determine the bandwidth for geographically weighted factor analysis estimation. For this purpose the authors developed a software implementation using the statistical framework R. The investigation of the accuracy of the criterion that determines the optimal number of neighbors is based on the results of the experiments. They show that proposed significance test statistics determines the optimal number of nearest neighbors more accurately. Furthermore the dependence of significance test statistics on the number of nearest neighbors appears smoother. This makes it possible to develop more effective optimization algorithms for automatic bandwidth selection. The proposed approach is used for education monitoring problems.

Acknowledgements

This research has been supported by the Ministry of Education and Science of the Russian Federation as part of the state task (project No 2.7996.2017/BCh).

References

[1] Lloyd CD. Analysing population characteristics using geographically weighted principal components analysis: a case study of Northern Ireland in 2001. Computers Environment and Urban Systems 2010; 34: 389–399.
[2] Harris P, Brunsdon C, Charlton M. Geographically Weighted Principal Components Analysis. International Journal of Geographical Information Science 2011; 25(10): 1717–36.
[3] Abdi H, Williams LJ. Principal component analysis // Wiley interdisciplinary reviews: computational statistics 2010; 2(4): 433–459.
[4] Zhang G. Estimating standard errors in exploratory factor analysis. Multivariate behavioral research 2014; 49(4): 339–353.
[5] Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots // Journal of the American statistical association 1979; 74(368): 829–836.
[6] Brunsdon C, Fotheringham AS, Charlton ME. Geographically Weighted Regression: a Method for Exploring Spatial Nonstationarity. Geographical Analysis 1996; 28(4): 281–298.
[7] Farber S, Páez A. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. Journal of Geographical Systems 2007; 9(4): 371–396.
[8] Cho SH, Lambert DM, Chen Z. Geographically weighted regression bandwidth selection and spatial autocorrelation: an empirical example using Chinese agriculture data. Applied Economics Letters 2010; 17(8): 767–772.
[9] Guo L, Ma Z, Zhang L. Comparison of bandwidth selection in application of geographically weighted regression: a case study. Canadian Journal of Forest Research 2008; 38(9): 2526–2534.
[10] Xu QS, Liang YZ. Monte Carlo Cross Validation. Chemometrics and Intelligent Laboratory Systems 2001; 56(1): 1–11.
[11] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL: http://www.R-project.org/ (15.05.2017).
[12] Information and analytical materials on the results of monitoring the effectiveness of educational institutions of higher education. URL: http://indicators.miccedu.ru/monitoring/2015/ (15.05.2017).