Mathematical cartographic approaches toward evaluation and forecasting of ethnic composition in the Russian regions

Sergey Riazantsev, Vladimir Tikunov, & Sergey Timonin

a Institute of Socio-Political Research, Russian Academy of Sciences, Moscow, Russian Federation
b Faculty of Geography, Lomonosov Moscow State University, Moscow, Russian Federation
c National Research University, Higher School of Economics, Moscow, Russian Federation

To cite this article: Sergey Riazantsev, Vladimir Tikunov & Sergey Timonin (2013): Mathematical cartographic approaches toward evaluation and forecasting of ethnic composition in the Russian regions, Annals of GIS, 19:1, 17-25

To link to this article: http://dx.doi.org/10.1080/19475683.2012.758174
Mathematical cartographic approaches toward evaluation and forecasting of ethnic composition in the Russian regions

Sergey Riazantseva, Vladimir Tikunovb and Sergey Timoninc

aInstitute of Socio-Political Research, Russian Academy of Sciences, Moscow, Russian Federation; bFaculty of Geography, Lomonosov Moscow State University, Moscow, Russian Federation; cNational Research University, Higher School of Economics, Moscow, Russian Federation

(Received 19 November 2012; final version received 30 November 2012)

This article is devoted to the study of the population’s ethnic structure in regions of Russia and former RSFSR (Russian Soviet Federative Socialist Republic) as well as the temporal dynamics of major ethnic groups by means of mathematical and cartographic modelling. Integrated indicators are developed to estimate ethnic diversity in regions of Russia and former RSFSR (ethnic diversity index and its modification – ethnic diversity index adjusted for the ability to speak Russian), and cluster analysis is performed to offer typological classification of Russian regions based on their ethnic composition. Maps are created on the basis of the derived indicators and typological classification. Finally, the estimates of the share of major ethnic groups up to 2030 are provided.

Keywords: ethnic diversity index; typological classification; GIS-mapping; ethnic forecasting

Introduction

During the last couple of centuries, mankind has been divided into different groups according to their physical differences, cultural features, political peculiarities, etc. As a rule, such information is collected in the course of a population census and is based on people’s self-identification. On the one hand, the majority of countries in the world regularly carry out censuses and this procedure has become more or less routine and conventional. But on the other hand, there exists a lot of differences (including questionnaires) and socio-political consequences.

In the introduction of this paper, we are not going to provide the reader with a detailed anthropological explanation of differences between the below-mentioned terms, but give a brief overview of those categories that can be collected during the censuses with precise consideration of the censuses in Russia and former Soviet Union. It is also important to stress that we consider categories of people exclusively for carrying out demographic research.

Let us start with citizenship, because such information is usually collected in all censuses. Citizenship is a term that is sometimes used to denote the status of those nationals who have full political privileges. According to the Principles and Recommendations for Population and Housing Census (2008), information on citizenship can be collected so as to permit the classification of the population into three categories: (1) citizens by birth; (2) citizens by naturalization, whether by declaration, option, marriage, or other means; and (3) foreigners. In addition, information about the country of citizenship of foreigners can be collected. The last information is of special importance for countries with a huge migratory inflow such as the United States, Canada, Australia, Great Britain, and Russia.

Some countries also collect information on ‘race’ and/or ‘ethnicity.’ The decision to collect and disseminate such information in a census is dependent upon a number of considerations and national circumstances, including, for example, the national needs for such data, and the suitability and sensitivity of asking ethnicity questions in a country’s census. Identification of the ethnocultural characteristics of a country’s population has an increasing importance in the context of migration, integration, and policies affecting minority groups (Tishkov 2011).

Race, as a categorizing term referring to human beings, was first used in the English language in the late sixteenth century. Since then, it has had a variety of meanings. What most definitions have in common is an attempt to categorize people primarily by their physical differences: skin color, hair texture, facial features, and eye formation (Long and Kittles 2003). For example, in the first US census in 1790, two races were considered: white and others. Almost two hundred years later, in 1977, the United States made an
National identity
Citizenship
Native language
Ability to speak Russian
Ability to speak other languages
Place of birth
Duration of residence in the current place of living
Religion affiliation

number of nationalities cannot be treated as an objective indicator of the process of ethnic diversity growth in Russia.

Issues of ethnicity as well as its impact on socioeconomic development of the country have been examined by numerous researchers from a variety of academic disciplines. The main problem of our research is the investigation of ethnic structure in Russia within space and time (spatial patterns and temporal dynamics). We strongly believe that ethnic composition of the population and its changes correlates with the demographic processes and this relation is going to be the topic of our future research.

The research objective includes the creation of a complex indicator that allows us to make a comparison of Russian regions on ethnic structure between each other and over time. In this connection, we introduce the ethnic diversity index (EDI) and its version – EDI adjusted for changes over time. In this connection, we introduce the ethnic diversity index (EDI) and its version – EDI adjusted for changes over time. We strongly believe that ethnic composition of the population and its changes correlates with the demographic processes and this relation is going to be the topic of our future research.

The research objective includes the creation of a complex indicator that allows us to make a comparison of Russian regions on ethnic structure between each other and over time. In this connection, we introduce the ethnic diversity index (EDI) and its version – EDI adjusted for changes over time. We strongly believe that ethnic composition of the population and its changes correlates with the demographic processes and this relation is going to be the topic of our future research.

Estimation techniques of ethnic diversity in Russian regions

In order to estimate how ethnically diverse populations across Russian regions are, we modified the technique developed by Phil Meyer and Shawn McIntosh for the United States (Philip and McIntosh 1992). They created USA TODAY Diversity Index that measures the probability that two persons, selected at random from the same census area (region), belong to two different races or ethnic groups. The index uses two basic principles of probability theory: (1) to obtain the probability that all of several independent events will occur, multiply their separate probabilities and (2) to obtain the probability that at least one of several independent events will occur, add their separate probabilities. The formula for calculations is as follows:

\[
\text{Diversity} = 1 - \left( (W^2 + B^2 + \text{AmInd}^2 + A^2 + \text{NH}^2) / (H^2 + \text{Non} - H^2) \right) \tag{1}
\]

where, in the first part of the equation, squared proportions of each race counted by the federal government are summed, and in the second part, squared proportions of each ethnicity are added. Two parts of the equation are multiplied and subtracted from 1. W means white, B means black, AmInd means American Indian, A means Asian, NH means native Hawaiian, H means Hispanic, and Non-H means non-Hispanic.

Applying this idea to our research, we have created EDI, and the algorithm of its computation is described below. At first, we calculate the share of each ethnic group in every region that should be interpreted as the probability that randomly chosen person is of a certain ethnicity:

\[
P_{ij} = \frac{\text{Pop}_{i,j}}{\sum_{j=1}^{m} \text{Pop}_i} \tag{2}
\]

where \( j \) is the region of Russia or RSFSR, \( i \) is the ethnic group in a region, \( m \) is the number of regions of interest (total number of regions in the Russian Federation or RSFSR), \( n \) is the number of ethnic groups registered in a region, \( \text{Pop}_{i,j} \) is the population size of the \( i \)th ethnic group in the \( j \)th region, and \( \sum_{j=1}^{m} \text{Pop}_i \) is the population size of the \( i \)th ethnic group in Russia or RSFSR. After the described normalization, results can be presented in a matrix form:

\[
\begin{bmatrix}
P_{1,1} & P_{1,2} & \ldots & P_{1,n} \\
P_{2,1} & P_{2,2} & \ldots & P_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
P_{m,1} & P_{m,2} & \ldots & P_{m,n}
\end{bmatrix} \tag{3}
\]

Second, we square that probability (or multiply single probability by itself). This case already describes the probability that two persons, drawn at random, will be from that particular ethnic group. Third, we sum the squared probabilities for each region and get the final probability that two randomly selected people are from the same ethnic group (within each region). Finally, we subtract this derived probability from 1 to get the probability that two random people belong to different ethnic groups and multiply by 100 to get an integer:

\[
\text{EDI}_j = \left( 1 - \sum_{i=1}^{n} P_{ij}^2 \right) \times 100 \tag{4}
\]

Index value may range from 0 (no diversity, ethnically homogenous region) to 100 (complete diversity, ethnically diverse region). The value ‘0’ means that only one ethnic group is presented in a region, whereas the index point ‘100’ can be interpreted as if two individuals, chosen at random, have zero probability to be of the same ethnicity.

We have implemented our methodology to compute EDI for regions of Russia and RSFSR based on the 1970,
1979, 1989, and 2002 Census data. The national indices (calculated for the whole Russia or RSFSR) are 31.1 (1970), 31.5 (1979), 33.2 (1989), and 36.1 (2002). One can notice that the national EDI has been constantly increasing since the 1970 Census (Figure 1), but the steepest growth was observed after the USSR collapse (between the 1989 and 2002 Census).

In Table 2, 20 Russian regions with the calculated EDI are presented (according to the 2002 Census) (Figure 2). In the first columns, information about 10 most ethnically diverse Russian regions is accumulated. This group of regions is fully composed of either ethnic republics or autonomous districts. In the second half of Table 2, ethnically homogeneous Russian regions are presented.

Features 1 and 2 reflect the distribution of EDI across Russian regions and over time. According to our calculations, Russian regions vary considerably in their ethnic composition. Besides this, comparison of two maps allows us to conclude that ethnic diversity has increased significantly between 1970 and 2002 but in various regions differently. Chukotka Autonomous District, Republic of Bashkortostan, Khanty-Mansi Autonomous District – Yugra, Stavropol Territory, and Koryakiya Autonomous District have become more diverse from the ethnic point of view. Meanwhile, Udmurt Republic, Republic of Tyva, Republic of Kareliya, and Jewish Autonomous Region have become more ethnically homogenous (Figure 3).

We also made an attempt to modify the EDI by including additional information about the share of people speaking/not speaking the Russian language as the only official language in Russia (Timonin 2010). The algorithm of our calculations is as follows:

<table>
<thead>
<tr>
<th>Number</th>
<th>Name of the Russian regions</th>
<th>EDI</th>
<th>Number</th>
<th>Name of the Russian regions</th>
<th>EDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Republic of Dagestan</td>
<td>83.8</td>
<td>80</td>
<td>Yaroslavl Region</td>
<td>9.4</td>
</tr>
<tr>
<td>2</td>
<td>Republic of Bashkortostan</td>
<td>72.0</td>
<td>81</td>
<td>Tula Region</td>
<td>9.3</td>
</tr>
<tr>
<td>3</td>
<td>Karachayev-Cherkesskaya Republic</td>
<td>71.9</td>
<td>82</td>
<td>Arkhangelsk Region</td>
<td>9.3</td>
</tr>
<tr>
<td>4</td>
<td>Chukotka Autonomous District</td>
<td>66.5</td>
<td>83</td>
<td>Oryol Region</td>
<td>9.1</td>
</tr>
<tr>
<td>5</td>
<td>Koryakiya Autonomous District</td>
<td>66.5</td>
<td>84</td>
<td>Kostroma Region</td>
<td>8.6</td>
</tr>
<tr>
<td>6</td>
<td>Yamalo-Nenets Autonomous District</td>
<td>62.9</td>
<td>85</td>
<td>Lipetsk Region</td>
<td>8.2</td>
</tr>
<tr>
<td>7</td>
<td>Taymyr Autonomous District</td>
<td>62.7</td>
<td>86</td>
<td>Kursk Region</td>
<td>8.1</td>
</tr>
<tr>
<td>8</td>
<td>Republic of Sakha (Yakutia)</td>
<td>62.1</td>
<td>87</td>
<td>Bryansk Region</td>
<td>7.2</td>
</tr>
<tr>
<td>9</td>
<td>Kabardino-Balkarskaya Republic</td>
<td>61.7</td>
<td>88</td>
<td>Tambov Region</td>
<td>6.9</td>
</tr>
<tr>
<td>10</td>
<td>Republic of Kalmykiya</td>
<td>60.1</td>
<td>89</td>
<td>Vologoda Region</td>
<td>6.7</td>
</tr>
</tbody>
</table>
Figure 2. Ethnic diversity in regions of Russia (EDI is computed on 2002 Census data).

Figure 3. Dynamics of ethnic diversity in 1970–2002, abs.

(1) The probability that someone speaks Russian or not is computed separately. Thus, we calculate the share of Russian-speaking population and non-Russian-speaking population (in the whole region), square these two derived values, and sum them up.
(2) Multiply the sum of the ethnic probabilities and ‘Russian language’ probabilities. This is the probability that any two random persons are the same by ethnicity and both speak or do not speak Russian.

(3) Subtract this probability from 1 to get the probability that two random persons are different (of different ethnicity, and one speaks Russian while the other one does not). Multiply by 100 to get an integer:

$$\text{EDI}_{Rj} = \left(1 - \left(\sum_{i=1}^{n} P_{ij} \times \left(\text{Rus}_j^2 + \text{Non} - \text{Rus}_j^2\right)\right)\right) \times 100$$

where $\text{Rus}_j^2$ is the share of people in the $j$th region who chose Russian as the main language during the census and $\text{Non} - \text{Rus}_j^2$ is the share of people in the $j$th region who chose language different from Russian as the main.

By adding information about the ability to speak the Russian language, we have the opportunity to describe population diversity more precisely. In Table 3, 10 Russian regions are presented with the largest absolute difference between EDI and EDI adjusted for the ability to speak Russian. For example, calculations based on two methodologies have the most distinguished results for the Chechen Republic. If we appeal to the initial data, we can see that the share of Chechens in the Chechen Republic is more than 93% that makes this region almost ethnically homogenous. However, the proportion of those who speak Russian is also big enough (81.4%) which increases ethnolinguistic diversity in the Chechen Republic.

### Typology of Russian regions on the population ethnic composition

The next step of our research was to develop the typology of Russian region on the ethnic composition. For these purposes, we used the algorithm of cluster analysis proposed elsewhere by the present author (Tikunov 1997). The share of five major ethnic groups in a region was chosen as basic indicators. The offered method assumes normalization of initial indicators on dispersions after what the matrix identical to a matrix (3) is formed. The largest Euclidean distance gets out of this matrix and two regions connected by this distance become kernels and constitute homogeneous groups (taxons). These groups are formed by distribution of remained $(n - 2)$ regions between the two kernels on the minimum of Euclidean distances. In this case, both groups are created under the condition of minimization of within-group differences.

At the second stage (formation of three groups), the algorithm works as follows. The two first kernels remain, and the third kernel is being searched according to the following scheme. Each of remained $(n - 2)$ regions is tested as the third kernel, and remained $(n - 3)$ regions are distributed between three kernels under condition of minimization of Euclidean distances. For each variant of grouping, the sum of within-group differences is calculated, and that variant which gives the smallest sum is accepted as the final one for three-group division. The algorithm works similarly before the formation of necessary number of homogeneous groups.

During the simulation, we tested 4–10 groups. Finally, six types/clusters of regions were defined from the substantial point of view (Figure 4).

The first type of areas unites the regions with strongly pronounced unipolar ethnic structure. Only one ethnic group obviously dominates in each of these regions; all the other ethnic groups constitute a small share. Many Russian-speaking regions of European Russia, Siberia, and the Volga region make up this group. These are regions mainly populated by the ethnic Russians: St.-Petersburg, Pskov, Leningrad, Vologda, Yaroslavl, Kostroma, Vladimir, Kurgan, Tomsk, Novosibirsk, Kemerovo, Irkutsk areas, and some other regions.

The second group is formed by the regions with strongly pronounced unipolar structure, but with the allocated second ethnos. City Moscow; Moscow, Tula, Oryol, Belgorod, Kostroma, Sverdlovsk, and Amur areas; Khabarovsk and Primorsk territories; and Jewish autonomous district constitute this group.

<table>
<thead>
<tr>
<th>Number</th>
<th>Name of Russian regions</th>
<th>EDI</th>
<th>EDI$_R$</th>
<th>EDI$_R$ − EDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chechen Republic</td>
<td>12.5</td>
<td>39.0</td>
<td>26.5</td>
</tr>
<tr>
<td>2</td>
<td>Republic of Ingushetiya</td>
<td>36.1</td>
<td>50.9</td>
<td>14.8</td>
</tr>
<tr>
<td>3</td>
<td>Republic of Tyva</td>
<td>36.6</td>
<td>50.2</td>
<td>13.6</td>
</tr>
<tr>
<td>4</td>
<td>Saint-Petersburg</td>
<td>28.1</td>
<td>35.9</td>
<td>7.8</td>
</tr>
<tr>
<td>5</td>
<td>Moscow</td>
<td>27.9</td>
<td>32.9</td>
<td>5.0</td>
</tr>
<tr>
<td>6</td>
<td>Aginsk Buryatsk Autonomous District</td>
<td>48.6</td>
<td>53.5</td>
<td>5.0</td>
</tr>
<tr>
<td>7</td>
<td>Ivanovo Region</td>
<td>12.2</td>
<td>17.1</td>
<td>4.9</td>
</tr>
<tr>
<td>8</td>
<td>Republic of Sakha (Yakutiya)</td>
<td>62.1</td>
<td>66.9</td>
<td>4.8</td>
</tr>
<tr>
<td>9</td>
<td>Moscow Region</td>
<td>17.1</td>
<td>21.1</td>
<td>4.0</td>
</tr>
<tr>
<td>10</td>
<td>Republic of Tatarstan</td>
<td>56.3</td>
<td>59.9</td>
<td>3.7</td>
</tr>
</tbody>
</table>
The third cluster includes territories with expressed unipolar ethnic structure, but with the presence of other ethnicities. Russians constitute the basic part of the population in these regions, but the presence of other people is not simply considerable, but quite essential. Krasnodar, Perm, Krasnoyarsk, and Stavropol territories; Rostov, Arkhangelsk, Volgograd, Orenburg, Samara, Tyumen, and Magadan areas; and Nenets, Yamal-Nenets, Hunts-Mansijsky, and Tajmyrsky autonomous districts make up this group.

Regions included in the fourth group are with unipolar ethnic structure with a considerable share of the second ethnos (conditionally bipolar structure). For example, these regions are Adygea, Kalmykia, North Ossetia, Ingushetia, Chuvashiya, Mordovia, Udmurtya, Komi, Altai, Khakassia, Tuva, Buryatiya republics; Chukchi, Evenki, and Koryak autonomous districts. As a rule, the second largest ethnic group in these regions are Russians.

The regions included in the fifth group are with ‘pure’ bipolar ethnic structure of the population (the size of two ethnic groups are almost equal and the share of these two groups is necessary more than 80% of the total population of region). Three regions can be carried into this group: Mary El, Tatarstan, and Yakutia republics. As a rule, Russians are the largest second ethnic group in such areas.

The regions included in the sixth group are the territories with multipolar ethnic structure of the population. A great number of different ethnic groups are presented in each region included in this cluster. No one ethnicity makes the obvious majority. Only four territories form this group: Bashkortostan, Dagestan, Kabardino-Balkariya, and Karachaevo-Circassia republics. With reference to these territories, the national policy of ethnic quota system in different spheres of public life is required.

Medium-term forecast of the ethnic structure of the population in Russia

The questions of population forecasting by ethnic groups are discussed in several papers (Large and Ghosh 2006a; Coleman 2010, etc.). In most of the cases, cohort-component methodology is applied that requires basic data on initial age–sex structure of the populations, on fertility, mortality, and migration of the ethnic groups being projected as well as assumptions about their future levels and trends. Since there is no complete source of information about age-specific fertility and mortality rates of the ethnic groups in Russia, we are unable to use cohort-component method for ethnic population projections in full measure. Besides, this it is not the main objective of the article – just to have a look at the possible ethnic structure of Russia in 2030. In this connection, we use simple mathematical algorithm to try to estimate future size of the major ethnicities in Russia. Beyond any doubt, the future trends of ethnic dynamics are in fact quite uncertain. They will depend on both natural increase of various ethnic groups and ethnic
Table 4. Change rates of the major ethnic groups in the intercensus periods.

<table>
<thead>
<tr>
<th>Annual growth rates for each intercensus period (t years) (%)</th>
<th>Average from the annual growth rates for four intercensus periods (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population</td>
<td>0.61</td>
</tr>
<tr>
<td>Russians</td>
<td>0.58</td>
</tr>
<tr>
<td>Tatars</td>
<td>0.57</td>
</tr>
<tr>
<td>Ukrainians</td>
<td>0.99</td>
</tr>
<tr>
<td>Bashkirs</td>
<td>1.00</td>
</tr>
<tr>
<td>Chuvashs</td>
<td>0.35</td>
</tr>
<tr>
<td>Armenians</td>
<td>2.46</td>
</tr>
<tr>
<td>Others</td>
<td>0.64</td>
</tr>
</tbody>
</table>

features of migratory processes. To a first approximation, we just applied average annual growth rate derived from retrospective data to forecast the share of major ethnic groups in Russia up to 2030.

On the first step, it is necessary to define tendencies and patterns of transformation of ethnic composition during the previous time periods. For this purpose, calculations of dynamics of basic ethnic groups in the intercensus periods were performed. First, we simply computed annual growth rate of each major ethnic group for each intercensus period (1970–1979, 1979–1989, 1989–2002, and 2002–2010) according the following formula:

\[
GR_{[t_0; t]} = \left( \sqrt[\frac{t}{\text{Period}}]{\frac{\text{Pop}_t}{\text{Pop}_{t_0}}} - 1 \right)
\]


Figure 5. Medium-term forecast of the share of major ethnic groups in Russia up to 2030 (%). Color of labels reflects the trends of the change of the major ethnic group share (2010–2030): green – positive; red – negative; brown – no change.

Ukrainians (from 1.3% in 2010 to 0.9% in 2030) and Chuvashs (from 1.0% in 2010 to 0.9% in 2030). The share of the other people (except mentioned above) will probably grow significantly in Russia from 13.3% in 2010 to 16.6% in 2030 during the predicted period (Figure 5).

Conclusions

The Russian Federation is a multinational country that is written down at the very beginning of the national Constitution (it is almost a unique precedent in the world). Besides this, the country’s administrative-territorial division also underlines ethnic features – borders of some regions in Russia coincide with the areas of compact residing of ethnic groups. Such an approach was established in the Soviet period, but during recent times it has become a ‘delayed-action bomb’ for territorial integrity of the country. For instance, the Tatarstan, the Chechen Republic, and some other national regions have tried to proclaim their independence. For such a multinational state like Russia is, it is not only an interesting scientific problem to study
the ethnic structure of the population but also an important applied aspect since it can help to formulate the national and migratory policy.

Acknowledgment
This research is supported by the Russian Fund for Basic Research, project no. 12-06-00310 and no. 11-06-90413.

Notes
1. Before 1997, information about national identity was reflected in the passport. So, everyone should have presented information of his or her nationality.
2. It shares the official status at regional level with other languages in the numerous ethnic autonomies within Russia.

References


