RECOVERY OF DEMOGRAPHIC DATA OF RUSSIAN REGIONS BY POLYNOMIAL INTERPOLATION METHOD

Anzhelika Voroshilova – Mikhail Takmakov – Jeff L. Wafubwa

Abstract

The purpose of this work was to develop a technique allowing recovering of demographical data for various regions of the Russian Federation by applying linear and nonlinear polynomial interpolation. Statistical data on the population of the Russian Federation and the Russian Soviet Federative Socialist Republic for 1915-2020 was used as an empirical base for 5 regions. Missing statistics were restored using polynomial and spline interpolation. The data recovery method was verified by restoring a sparse and most complete data set for the Republic of Buryatia. Experimental results show that the proposed method has a high (1-3%) accuracy in the restoration of demographic data. Cluster analysis and mathematical statistics methods were applied to the restored demographic data for the regions. A method of comparative assessment of migration based on integro-differential analysis of population size and growth has been developed. In addition, it is shown that the use of certain methods of mathematical modeling in the framework of retrospective analysis can be their approbation for building demographic forecasts in the future.

Key words: regional demography, Russia, polynomial interpolation, data analysis.

JEL Code: C 610, C 650, C 670.

Introduction

There are multiple back-projection and inverse projection studies of methods of reconstructing population data for a specific historical period (Oeppen, 1993). For example, the investigation of the number of people who have ever been born (Cohen, 2014), the dynamics of the size of certain populations by age group (Sanchez-Romero et al. 2017) or territory of residence (Palma, Reis & Zhang, 2019; Edvinsson, 2015), etc.

From a methodological point of view, the restoration of demographic data began with numerical modeling methods in the framework of the theory of a stable population (Andreev & Volkov, 1977). However, with the development of numerical modeling methods, the range of modeling tools has expanded significantly and been refined (Keyfitz & Caswell, 2005; Bermúdez & Blanquero, 2016; Bonneuil, 2017).

Examples of such methods are: Lee’s inverse projection, back-projection methods, generalized inverse projection, trend projection, stochastic back-projection, Walle’s reconstruction method, Bonneuil reconstruction method, Bonneuil and Fursa’s reconstruction method, the reconstruction method by stochastic optimization, etc. (Bonneuil, 2017).

At the same time, there are no studies in the social sciences on the application of spline intepolation or polynomial interpolation methods for the purposes of inverse projection. Our study is an attempt to fill this gap and expand the range of application of new mathematical methods in the field of demographic inverse projection.

1 Data and methods

The purpose of our work was to reconstruct missing demographic data for various regions of the Russian Federation by applying the polynomial reconstruction method. Missing statistical data was restored using polynomial and spline interpolation. The data recovery method was verified by restoring a sparse data set for the Buryatia region.

Experimental results show that the proposed method has a high (1-3%) accuracy in restoring demographic data (Palma, Reis & Zhang, 2019). In this study we use an approach based on the polynomial interpolation method.

In general, the interpolation task is the process of restoring (completing) a function that is defined on a discrete set of points xi, i = 0, 1, . . . , n. It consists of the fact that given points with coordinates xi , yi where i=0, 1, 2, ... n are connected by straight line segments, and the function y (x) can be approximately represented as a broken curve (Habermann & Kindermann, 2007).

The cubic spline interpolation method creates a global fit. Global methods use all existing points to calculate coefficients for all interpolation intervals simultaneously. Accordingly, each data point affects the entire cubic spline. When moving one point, the entire spline changes accordingly, which makes it more uneven, and the more difficult it is to make it take the desired shape (Bermudez & Blanquero, 2016).

In our study we constructed df:

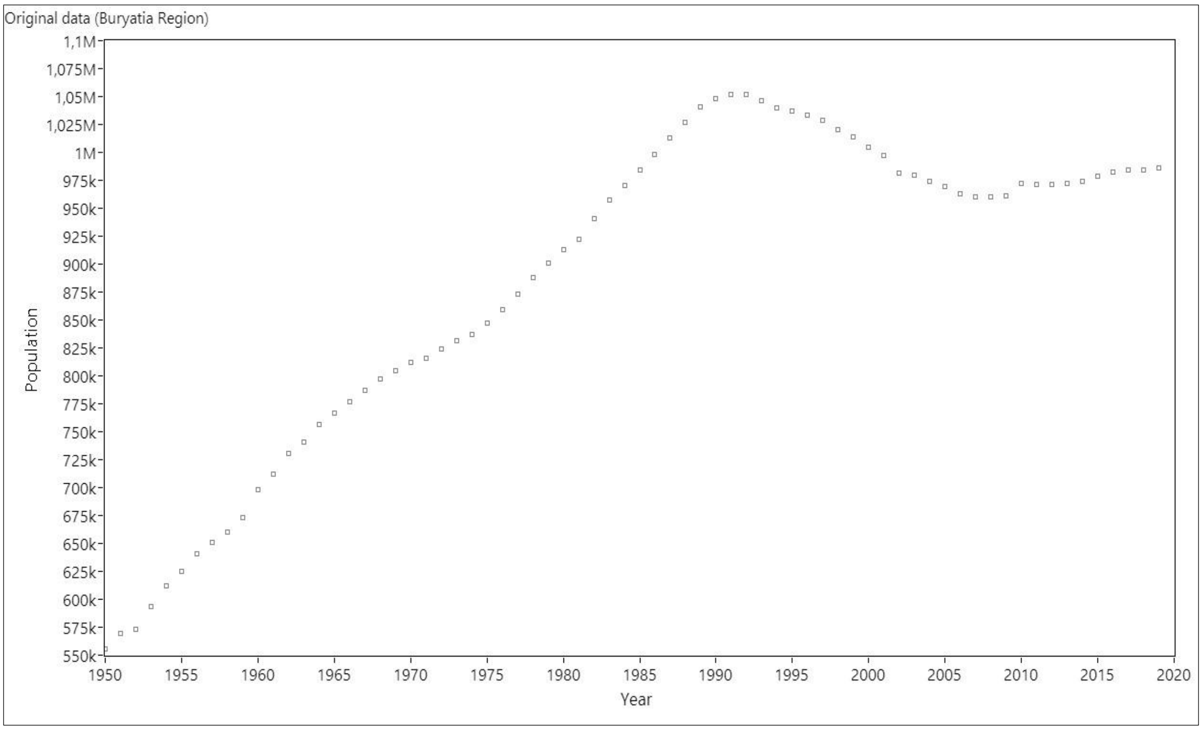
$$y = \sum_{i=0}^{N} x^i \times a\,[i],$$

where *y* is the population value and *x* is the year under consideration.

The result of applying this method is a nonlinear polynomial dynamic model, the coefficients of which are selected by the optimization method in accordance to the parameters stored in the algorithm (degree of the polynomial, permissible deviation, type of polynomial used, the method for estimating the error) (Gasca & Sauer, 2001).

To assess the performance and accuracy of the proposed approach, the most complete (without gaps) modal series of data were selected. For example, this criterion was met by a series of data reflecting the population of the Republic of Buryatia from 1950 to 2020 (Fig. 1).

Fig. 1: Population of the Republic of Buryatia (1950-2020)

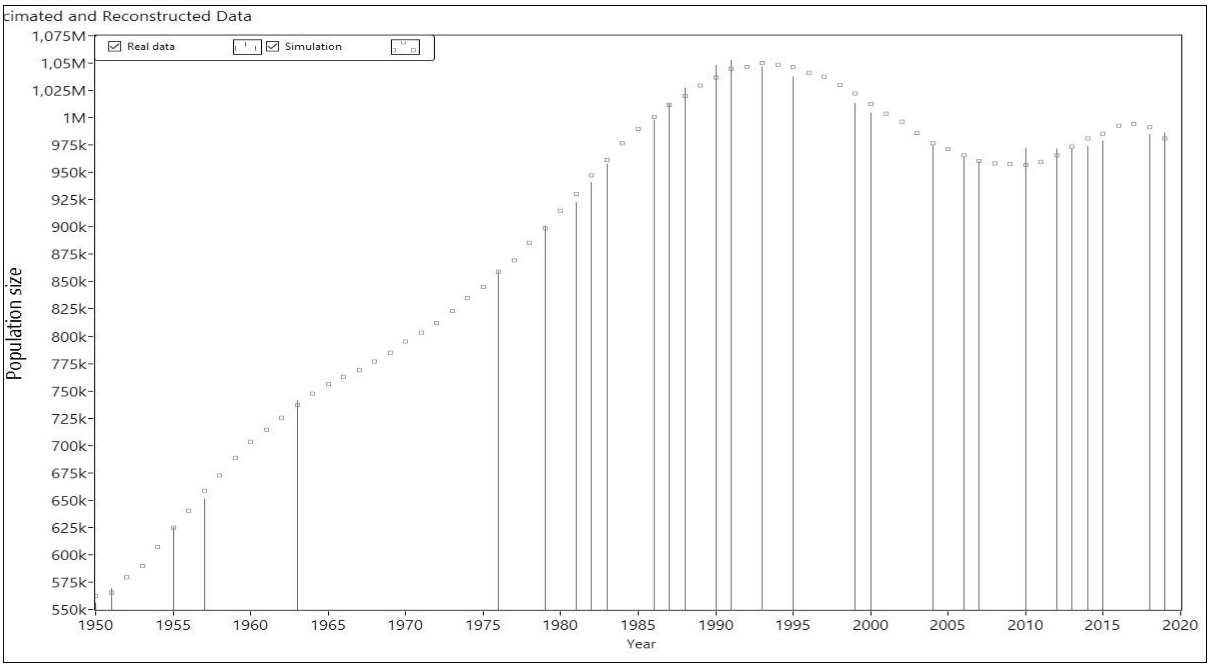


Source: Federal State Statistics Service

Furthermore, the modal row was decimated and 40% of the elements were randomly removed from it. We then built a polynomial model and restored the decimated data. As a result, the data was deleted in a way that resembled real cases of data loss.

For example in Fig. 2 one can see that the deleted data was absent both for short and long periods (wartime, for example). Figure 2 shows the decimated and restored data. One can see, the coincidence of the values restored by the polynomial interpolation method and real data is clearly visible. For a more complete assessment of the accuracy, the modeling error was calculated as the ratio of real data to model one in percent.

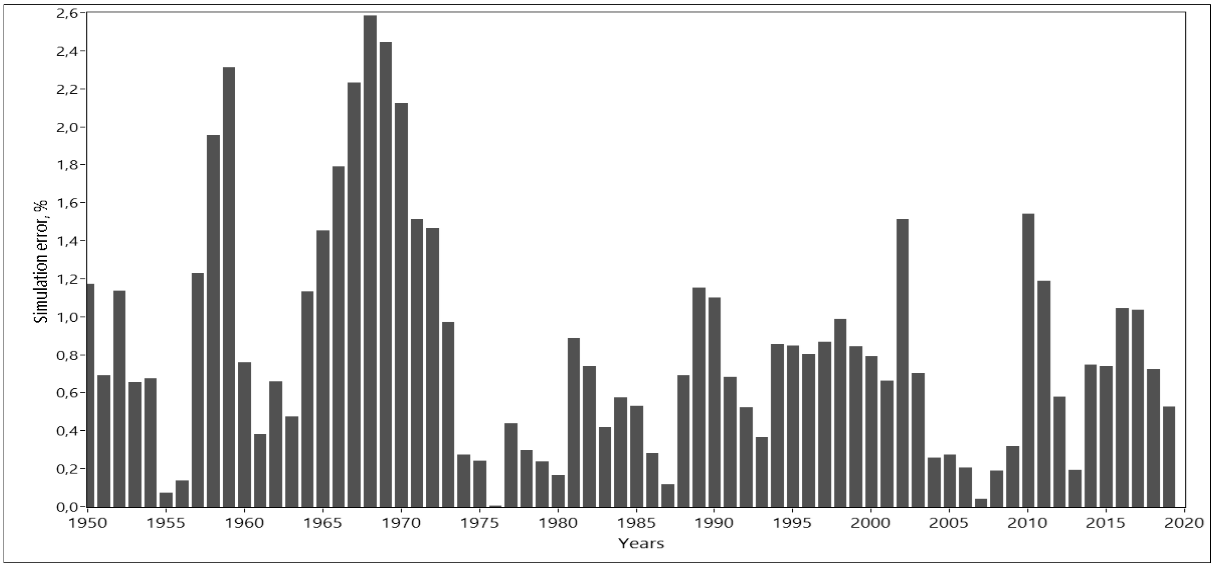
Fig. 2: The randomly decimated data of Buryatia region and restored data, obtained by polynomial model (1950-2020)



Source: Federal State Statistics Service and authors' calculation

Figure 3 shows that the error of the method as applied to this array does not exceed 2.6%.

Fig. 3: The error of restored data obtained by polynomial model, %



Source: the authors' calculation

We tested the proposed model on the most complete data of the Republic of Buryatia.

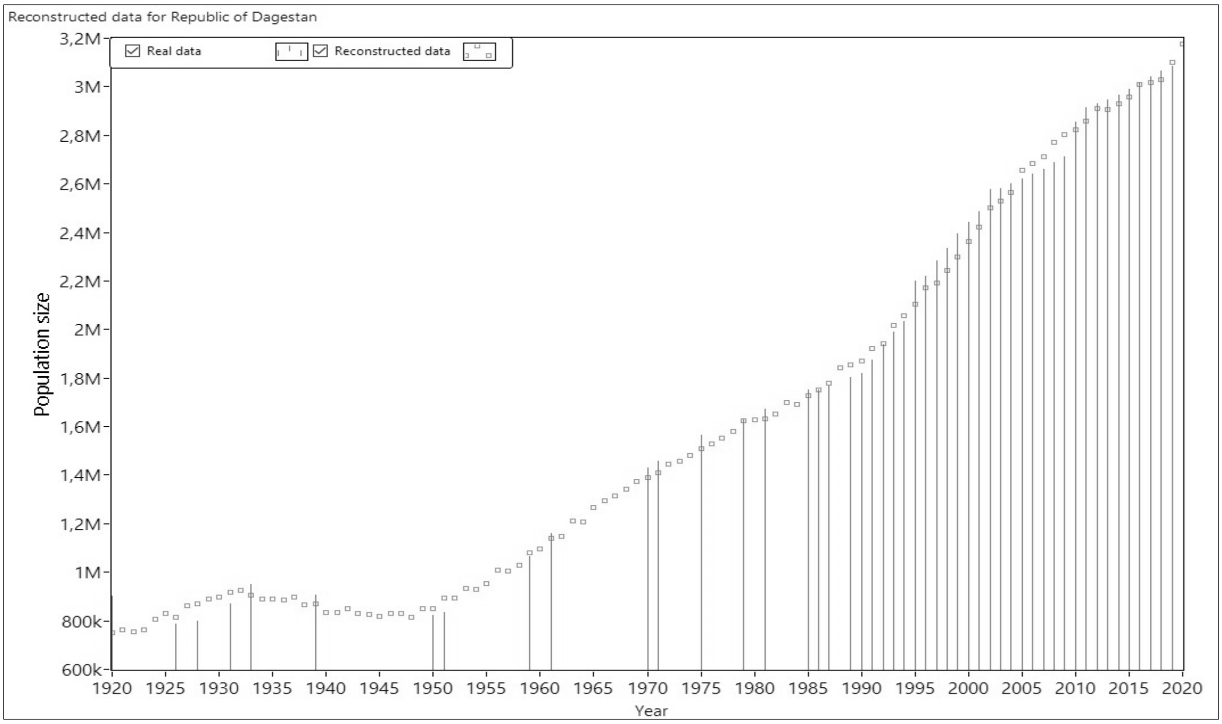
We then consistently applied the proposed approach to the data on the population in theother regions we selected. In particular, statistical data on the population of the Russian Federation and the Russian Soviet Federative Socialist Republic for 1915-2020 was used as an empirical base for 5 regions (Moscow, Leningrad, Irkutsk regions, as well as the republic of Dagestan).

2 Results

The data recovered by polynomial interpolation showed high accuracy of coincidence with real data in other regions of Russia. A perfect example of the demographic development in the Republic of Dagestan shows a steady population growth.

A slight decrease in the number is observed in the war and post-war years (1941-1950), which was not due to natural causes of population decline. At the same time, the interpolation polynomial model successfully demonstrates a high degree of accuracy in this segment, judging by partially available data (Fig. 4).

Fig. 4: The restored demographical data for Republic of Dagestan



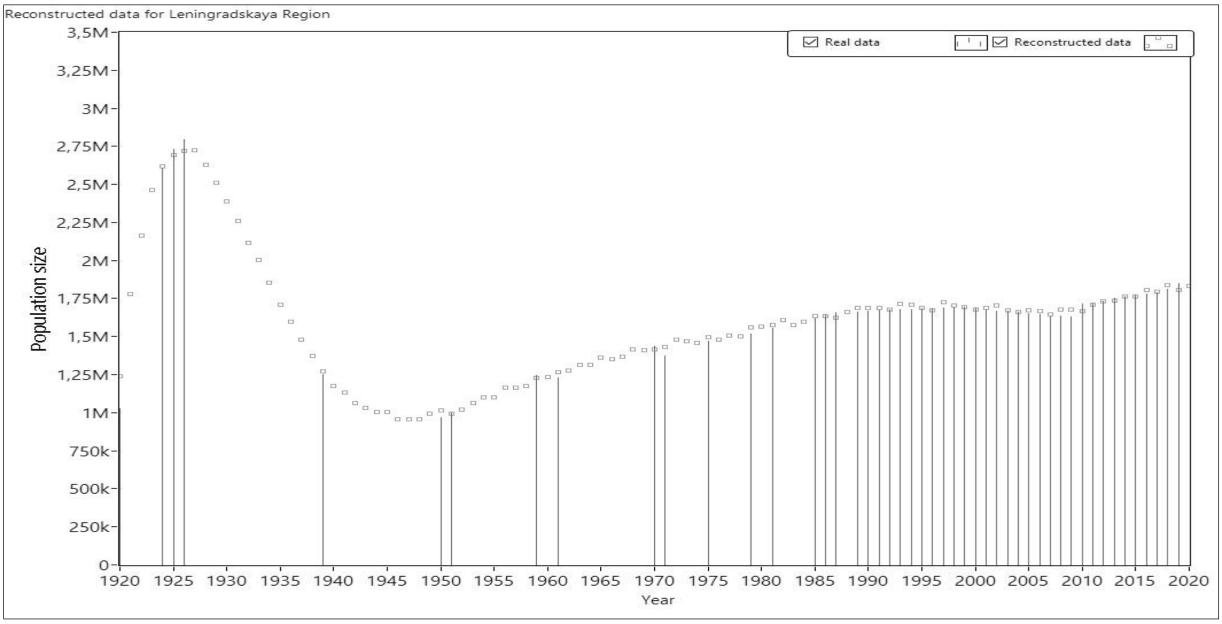
Source: Federal State Statistics Service and authors' calculation

Thus, the most significant result is that the polynomial interpolation method as a whole allows one to restore changes in demographic data caused not by “natural” reasons, but under the influence of social and cultural factors and events.

This is even more pronounced in the example of the Leningrad (Fig. 5) and Moscow Regions (Fig. 6).

The graph clearly depicts a sharp increase in the population of the central industrial regions before the war (industrialization of the 1930s), a sharp decrease in the population during the Second World War (1940s), and stable steady growth in recent years. Simulation data for these two regions also show a high level of relevance to the real situation.

Fig. 5: The restored demographical data for Leningrad Region

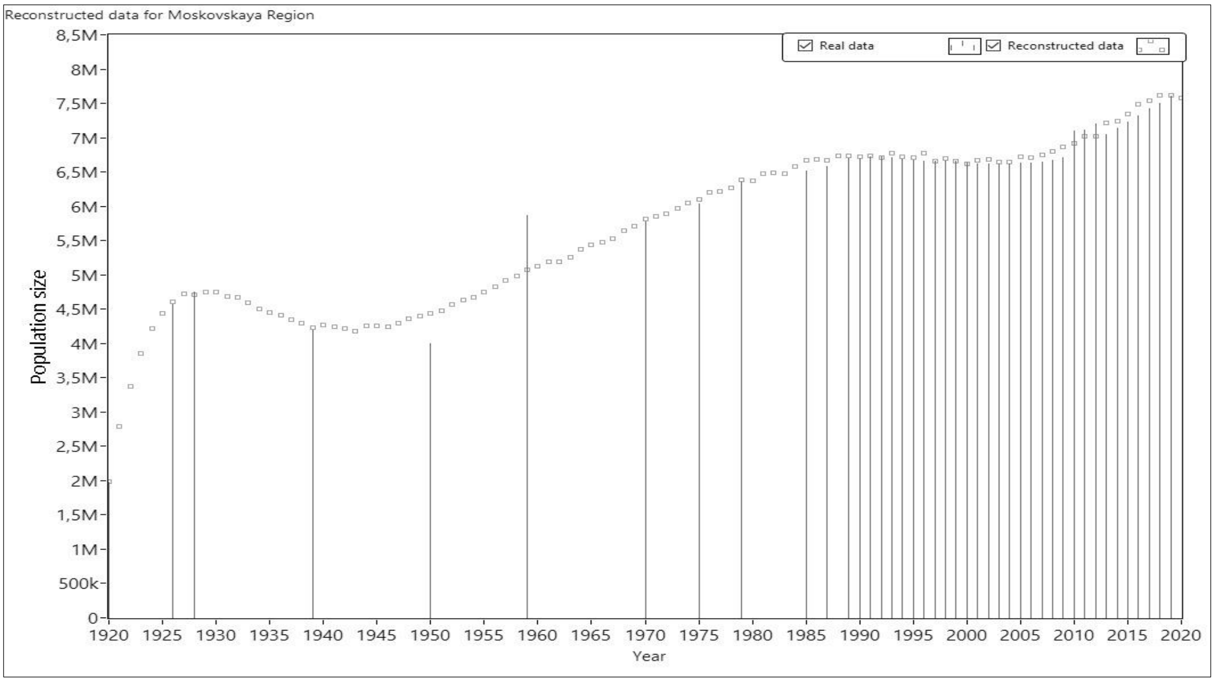


Source: Federal State Statistics Service and authors’ calculation

At the same time, figure 6 shows that the method imperfectly models the outburst of 1958. This shows the vulnerability of the polynomial interpolation method to single outliers at large void intervals.

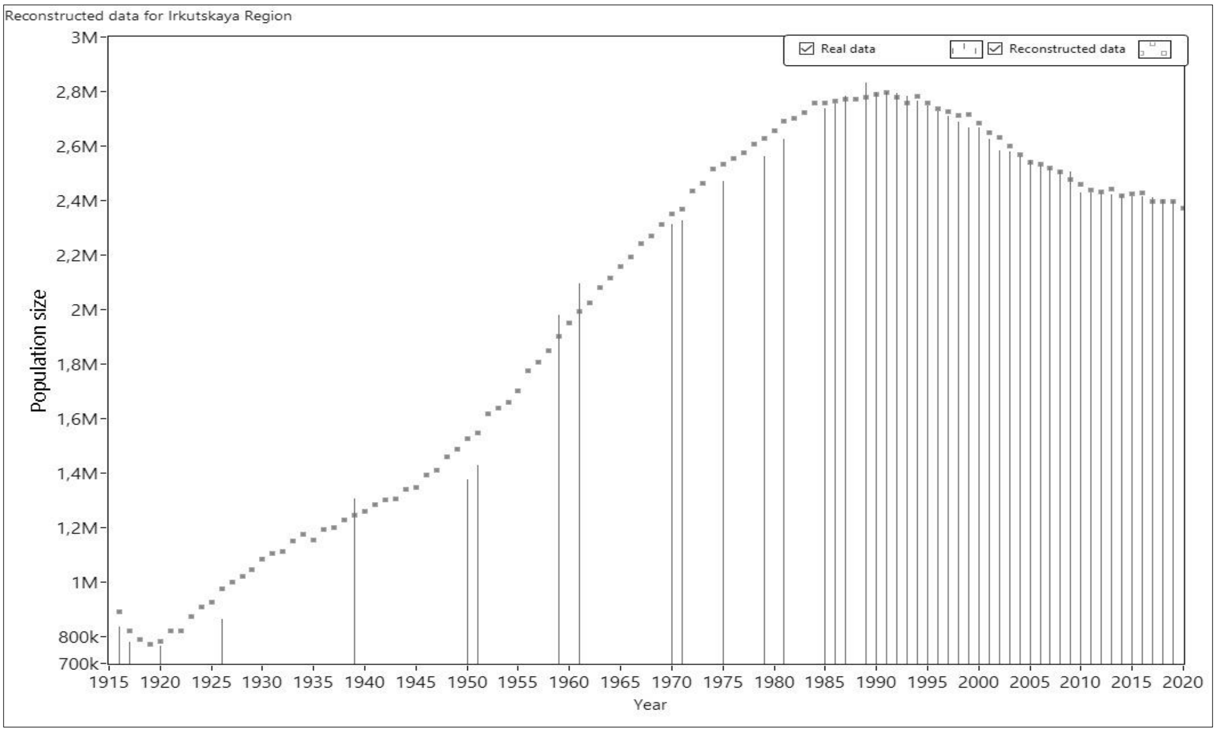
The abovementioned vulnerability of the method is confirmed by the example of applying polynomial interpolation to the restoration of lost demographic data on the population for the Irkutsk Region. The emissions of 1939, 1958, and 1961 were also not predicted correctly (Fig. 7).

Fig. 6: The restored demographical data for Moskow Region



Source: Federal State Statistics Service and authors' calculation

Fig. 7: The restored demographical data for Irkutsk Region



Source: Federal State Statistics Service e and authors' calculation

3 Discussion

Global demographic challenges of the last century such as the growth of the world's population as a whole, overpopulation in some regions and aging in others, demographic transition problems, global migration etc. have made demographic forecasting especially relevant. At the same time, it is obvious that modeling future demographic situations requires analysis of the demographic situation dynamics, based on past trends. It is possible to assess the impact of various factors on demographics by studying the statistical data of the past and comparing them with cultural and historical events in certain periods.

Thus, the complex use of numerical methods in combination with elements of sociological (socio-cultural) analysis makes it possible to identify significant socio-cultural determinants that influence the formation of the demographic situation to a high degree of probability. However, the problem is often that demographic data for certain past periods are incomplete, which does not allow for an in-depth analysis. A solution can be the use of modern methods of mathematical and computer modeling in the field of demography.

In this study, we considered the polynomial interpolation approach which showed that the dynamics of the population in the regions is due to a complex of intra-regional causes that deeply correlate with events of national scale (the Patriotic war, the Second World War, the crisis of the 1990s, etc.). In addition, it is shown that the use of certain methods of mathematical modeling in the framework of retrospective analysis can be their approbation for building demographic forecasts in the future.

Conclusion

The proposed interpolation polynomial model showed good functionality on demographic data. Thus, in our opinion, interpolation reconstruction can be useful for the partial recovery of descriptive demographic statistics for those periods, when it was not collected, reliable or lost. Moreover, the model well describes even those periods when population movement was not due to natural causes. At the same time, it was shown that the method remains vulnerable to single outliers over large void intervals of data. The advantage of this method is that it is possible not only to restore missing data but also to obtain fairly accurate demographic forecasts in the short term (5-10 years).

In our further studies we will develop a method of comparative assessment of migration based on integro-differential analysis of population size and growth rate.

Acknowledgment

The study was funded by a grant from the Russian Science Foundation (project no. 19-78-00058) “Analysis and simulation of demographical processes based on dynamical system approach with consideration of sociocultural determinants”.

References

Andreev, E. M., & Volkov, A. G. (1977). *Demograficheskie modeli*. Moscow: Statistika.

Bermudez, S., & Blanquero, R. (2016). Optimization models for degrouping population data. *Population Studies, 70(2),* 259–272. doi:10.1080/00324728.2016.1158853

Bonneuil, N. (2017). Reconstruction of populations by stochastic optimization: Sensitivity analysis. *Mathematical Population Studies*, *24*(3), 181–189. doi: 10.1080/08898480.2017.1330014.

Cohen, J. E. (2014). Is the fraction of people ever born who are currently alive rising or falling? *Demographic Research, 30(56),* 1561–1570. doi:10.4054/DemRes.2014.30.56.

Edvinsson, R. B. (2015) Recalculating Swedish pre-census demographic data: Was there acceleration in early modern population growth? // *Cliometrica, 9(2),* 167-191. doi: 10.1007/s11698-014-0112-z.

Federal State Statistics Service. (2020, January). Retrieved January 20, 2020, from https://www.gks.ru/bgd/regl/b18_14p/Main.htm.

Gasca, M., & Sauer, T. (2001). *Numerical Analysis: Historical Developments in the 20th Century*. Elsevier Science. doi: 10.1016/B978-0-444-50617-7.50007-0

Habermann, C., & Kindermann, F. (2007). Multidimensional Spline Interpolation: Theory and Applications. *Comput Econ, 30,* 153–169. doi: 10.1007/s10614-007-9092-4.

Keyfitz, N., & Caswell, H. (2005). *Applied Mathematical Demography* (3rd ed.). Springer. doi: 10.1007/b139042.

Oeppen, J. (1993). Back-Projection And Inverse Projection - Members Of A Wider Class Of Constrained Projection Models. *Population Studies-A Journal Of Demography*, *47*(2), 245–267. doi: 10.1080/0032472031000146996.

Palma, N., Reis, J., & Zhang, M. (2019). Reconstruction of regional and national population using intermittent census-type data: The case of Portugal, 1527–1864. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 53(1),* 11–27. doi:10.1080/01615440.2019.1666762.

Sanchez-Romero, M., Ediev, D., Feichtinger, G., & Prskawetz, A. (2017). How many old people have ever lived? *Demographic Research, 36*, 1667–1701. doi: 10.4054/DemRes.2017.36.54.

Contact

Anzhelika Voroshilova
Ural Federal University
620002, bld. 19, Mira st., Ekaterinburg, Russian Federation
a.i.voroshilova@urfu.ru

Takmakov Mikhail
Saint Petersburg Electrotechnical University "LETI"
197376, bld. 5, Professora Popova st., St. Petersburg, Russian Federation
mvtakmakov@stud.eltech.ru

Jeff L. Wafubwa
Saint Petersburg Electrotechnical University "LETI"
197376, bld. 5, Professora Popova st., St. Petersburg, Russian Federation
jefflusweti@gmail.com