Determinants of the Czech recent fertility transition - Spatial (geostatistical) modelling of a demographic process

Jaroslav KRAUS (Kraus@natur.cuni.cz)

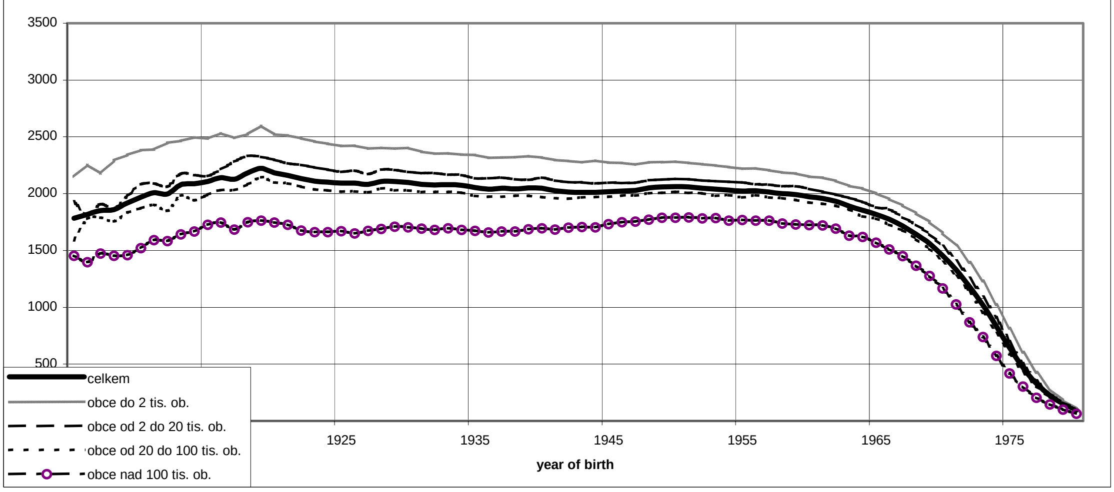
Charles University in Prague, Faculty of Science

Department of Demography andGeodemography

Keywords: Applications, Demography

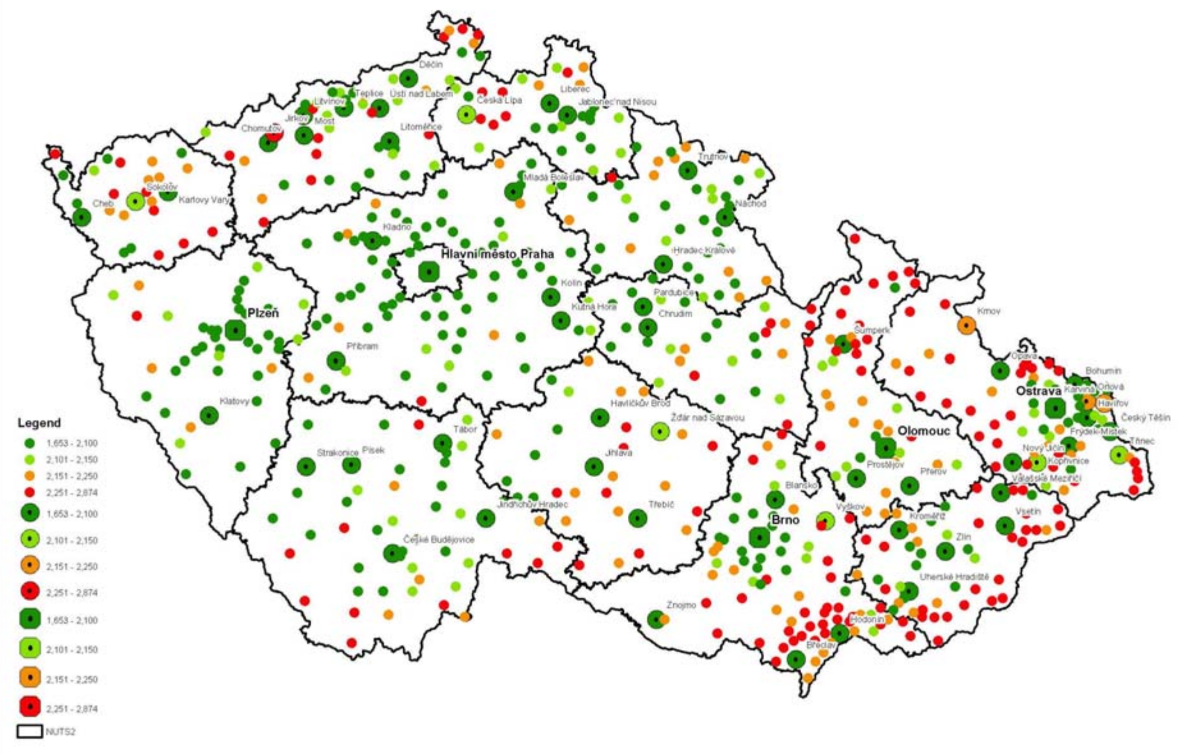
The data taken from population censuses are among the basic sources of demographic figures and they constitute the fundament of a number of analyses. Compared with others, they have the advantage of providing results in a major territorial detail, thus allowing, among others, an analysis of the data in terms of regional differentials. The definition of a territory is connected with the logical structure of the data to be analysed – in particular a certain minimum size of the sample of events which are the subject of the analysis. The basic demographic phenomena under observation certainly include *women's fertility*, while basic indicators of women's fertility include the generational (final) fertility of women. In the 2001 population census, all women aged 15 years and more replied to the question of how many live-born children they had. Along with the information about permanent residence, this data was a *source of analysis and model solution to the final fertility of women on the territory of the Czech Republic*.

Chart 1 Average number of children per 1,000 women in total, according to the size group of settlements



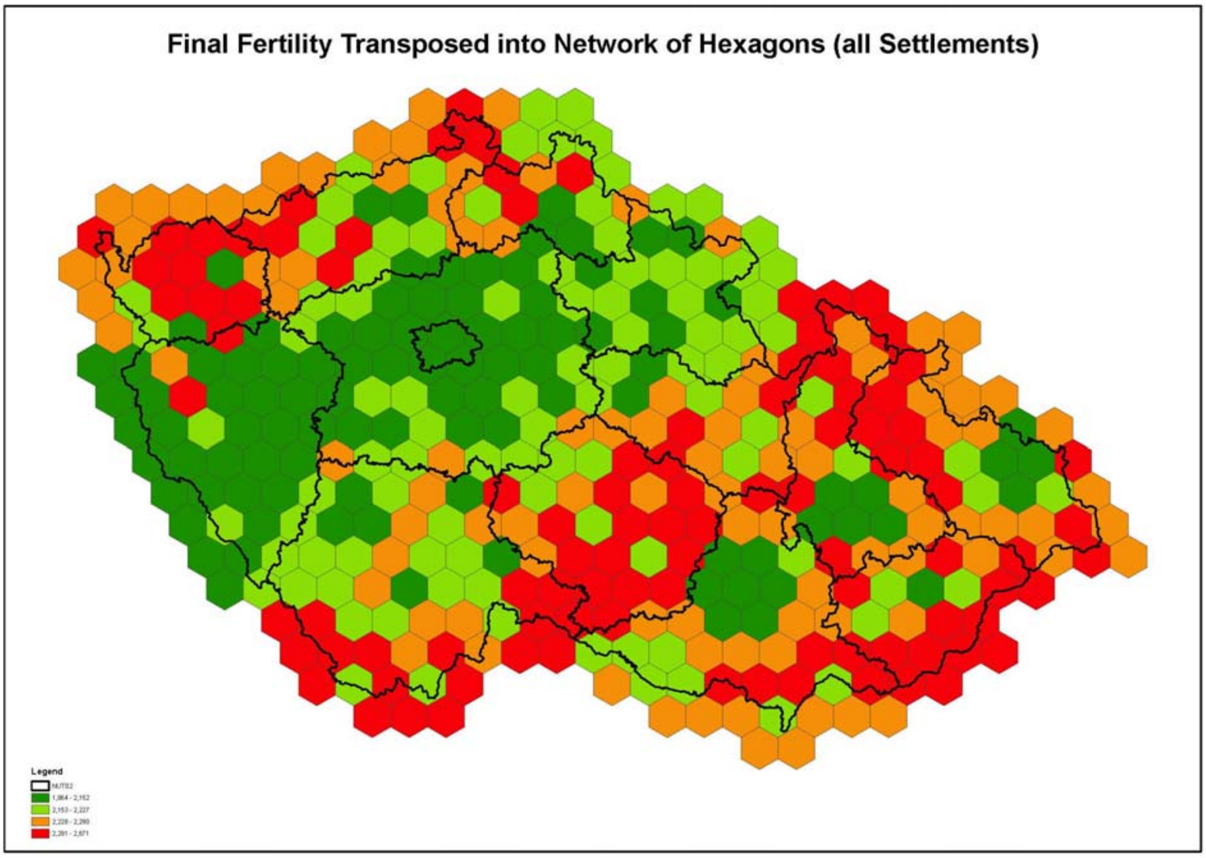
Women's final fertility can be expressed by the average number of living children. In the past decade, fertility dramatically plummeted in the Czech Republic. The total fertility rate fell from 1.89 in 1990 to the minimum of 1.13 in 1999 with a subsequent slight increase to 1.17 in 2002. With the trend, the Czech Republic's population has joined the category of lowest low fertility (Rychtaříková, 2004). As a result, the examination of changes in the fall has gained importance. Along with observation of the *causes of the fall*, there is the intriguing question of observation of the *changes from the territorial point of view*. A certain answer is provided by the results of the average number of children according to the regions and size groups of the place of residence. This has revealed (Kraus, 2002) that the average number of children is the highest in the smallest settlements and highest in large towns, especially in Prague. From the viewpoint of individual regions of the Czech Republic, the differences are not significant, with an obvious exception of Prague, where the average number of children is significantly lower.

1. Final Fertility by Settlements with more than 2000 Inhabitans



Now there is a new alternative which makes it possible to look at the spatial differentiation from a different angle – through *the geographical analysis of the Geographical Information System(GIS)*. In this case, the fundament is created by the geostatistical analysis which works with well-known data (points) in the space, creating (interpolating) a coherent area in the space above them. Within a geostatistical analysis, the value in every point of area is forecast from the measurements obtained in selected points. In general, there are two basic interpolation techniques: a determinist and a stochastic (geostatistical). Both methods rely on the similarity of phenomena in the close surroundings, while the geostatistical method adds the assumption of sampling error (ε) into the model of a territory. For a spatial analysis of demographic data, the latter approach is more convenient – after some basic problems are resolved. The first can be called *the ascertainment of values for a certain representative area* and it is connected with uneven character of settlement. This is clearly apparent on the first cartodiagram which depicts the average number of women in the settlements with more than 2,000 people. In this case, the problem is to calculate the women's fertility in the settlements with less than 2,000 people due to the size of the samples. For the purposes of this paper, the values in the settlements with less than 2,000 people were replaced with the value for all settlements of a given region (see the strata NUTS3), which is a certain simplification of the entire solution.

In order to achieve the representativeness of the area, one has to *construct the division of space into a regular network of cells (a mosaic), while the results for individual towns and villages are converted into the mosaic*. In practice, two mosaics are considered: one square and one hexagonal. The former has the advantage of being compatible with the structure of the data sequences used in computer technology and with the Cartesian system of coordinates. The latter (hexagonal) mosaic has the advantage that the centres of all neighbouring cells are equally distant from the centre of a given cell. Thanks to this symmetry, the hexagonal mosaic is methodologically more correct in terms of spatial analysis and this is why it was used in this paper – see cartodiagram 2. The total fertility of women in all settlements was converted into a network of hexagons with a 10-kilometre edge. The size of the edge of the hexagon was not selected randomly as it is connected with the average size of districts (NUST4), which are often used in demographic analyses.



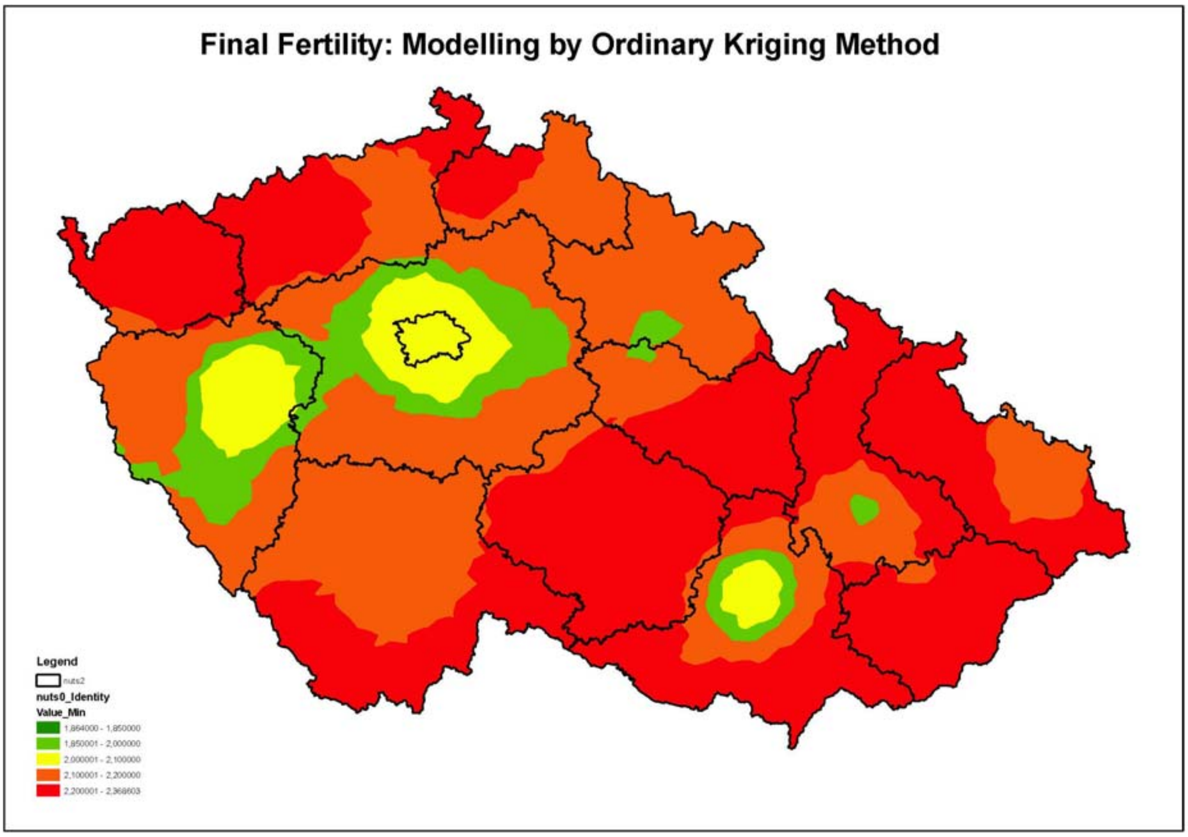
The calculation was based on individual (anonymous) samples of the census, in which the average number of children of all women (the women who had or had not children) was calculated in a standard way. The calculation of average number of children in the hexagonal mosaic was calculated as a weighted mean, with the number of women of relevant generation as the weight. In this way, the values of average number of children for 428 hexagons of the mosaic, which covers the territory of the Czech Republic, were calculated. In order to simplify the task, the Czech Republic was conceived as a closed entity, which implied that fertility beyond the Czech Republic´s border was zero. In practice, this meant enveloping the hexagonal tessalation by another edge beyond the Czech Republic´s border so that spatial functions converge to zero at the Czech Republic's border.

This created space for the work with the geostatistical model. When creating a model of a territory, data analysis is the first required step. Most, though not all, data models imply that data are quite or at least approximately normal. If the condition of normalcy of data is not fulfilled, the data must be transformed. However, since spatial data about the average number of children according to the territory are drawn from a roughly normal distribution, the data transformation is not necessary (see charts at cartodiagrams).

For the spatial solution, one of the Kriging methods was chosen. In fact, there is an array of them – both deterministic and stochastic. The Kriging methods are based on autocorrelation as a function of distance. In other words, the phenomena in a given locality are more influenced by phenomena in the neighbouring localities than in distant regions. The information about the value of a given phenomenon in a given space is calculated as distances between individual observations and the model autocorrelation as a function of distance: $Z(s) = \mu(s) + \varepsilon(s)$, where $Z(s)$ is the observed variable (the average number of children), split into the determinist trend $\mu(s)$ and random autocorrelated error $\varepsilon(s)$. The value s denotes the distribution in space: x, y coordinates of a given village and town. Like in other stochastic models, since there is the implication of imperfect forecasts of the determinist component, the value $\varepsilon(s)$ should on average amount to 0. Autocorrelation between $\varepsilon(s)$ and $\varepsilon(s+h)$ does not depend on real x,y values of s, but on the shift h between them.

When the task of this type is dealt with, another problem lies in the examination of the trend: if there is one (or if one can assume it a priori thanks to the knowledge of data), of what type it is – a constant or a linear function, which can be then expressed by a polynomial of an x-order. The Kriging methods are forecasting methods, with the forecast of an area under observation as the final objective (by means of various methods). It follows from this that it is possible to calculate the standardized error. As it ensues from the first, but even more from the second cartodiagram, one can discern a certain spatial pattern on the basis of the data about the average number of children: *the number of children in Western and Central part of the Czech Republic (Bohemia) is apparently lower than in the Eastern part (Moravia)*. As a result, the assumption of a trend can be built into the model, but it is problematic to build into it an estimate

$\mu(s)$. One can suggest replacing this value through the calculation of the average number of children for the entire Czech Republic. For these reasons, the method Universal Kriging was chosen, which assumes a continual trend for the data with an unknown value.



This paper had forecasting cartodiagrams carried out: *the depiction of the forecast value of the average number of children*. When creating the forecasting map, one has to work with a number of input parameters such as the number of neighbouring hexagons, which are taken into account in the calculation of the value in a given locality or the possible division of the total area into partial regions and the choice of the model function itself. At the end of the procedure, there is a cartodiagram as a model solution to the given task. The calculated value must be further specified: through a spatial enquiry, values for the territory of the Czech Republic must be obtained from the spatial solution expressed in the values of xmin, xmax, ymin and ymax. The results must be further subjected to a spatial scale – such as by means of quartile distribution of the phenomenon under observation and a bichrome colour scale. This results in a cartodiagram differentiating the territory into several contiguous territories: the area of Central and Eastern Bohemia with the lowest values, the region of Northern Moravia and the border region between Moravia and Bohemia in the south and a region with similar values in Western Bohemia.

To depict the chosen method, the alternative of the calculation of a probability map was drawn up, suggesting that the value in the given territory exceeds the defined constant. Two options can be chosen here: the value of average number of children for the entire Czech

Republic or the value of the average number of children of simple reproduction of 2.15. Again, it is necessary to work with a number of input parameters. A comparison of the model value with the replacement rate has a good descriptive ability: the Czech Republic is differentiated into three types of areas: Prague and its vicinity, where the value will not be reached, and the eastern part of the Czech Republic, where it is possible to accept the suggestion for the entire contiguous area.

In conclusion, one can state that geostatistical modelling of demographic phenomena does not principally differ from another sphere of demography – demographic forecasts. This is proved by the fact that the (geo)statistical apparatus is similar. The resulting solution is a forecast within a certain interval of reliability, burdened with a certain error. As such, one has to work with it. There is the undisputable advantage of clear plasticity of a geostastical forecast since its cartodiagrams are absolutely convenient for these purposes.

Bibliography:

Rychtaříková, J., Změny generační plodnosti v ČR se zaměřením na vzdělání žen (in Czech), Demografie 2, 2004

Kraus, J., Regionální diferenciace plodnosti (in Czech), Demografie 4, 2003

--------, Using ArcGIS Geostatistical Analyst, ESRI, 2001

Tuček, J., Geografické informační systémy, principy a praxe (in Czech), 1998, Computer Press