The KINA Model

A tool for exploring the spatial population development in China by large scale micro simulation

Peder Axensten[1], Zhongxin Chen[2], Shenghe Liu[3]

14th June 2004

Abstract: This paper is on work in progress. The KINA model exists and runs fast and usually reliably although it still has very simple behavioural modules that do not produce very reliable results. The purpose of the paper is to present the structure, content, and current state of validity of the model.

This is still work in progress. I hope to receive figures on nativity in, mortality in, emigration from, and immigration to China. It is of interest to compare these with the results from the model.

Background

Knowledge about the driving forces behind and the effects of the regional population development in China are imperative to a number of questions concerning the country's development and its relations to other countries. This includes questions such as ageing, the geographical distribution of people with various education levels, how the supply relates to the demand in the present and in the wanted development of trade and industry and how this development is influenced by regional efforts in education. Other questions concern land use, the restructuring of agriculture, migration and urbanisation, geographical distribution of income means etc.

A unique methodology to simulate, at micro level, the dynamics of large populations has been developed. The methodology builds on time geography, micro simulation, agent based systems and a special database technique. It was used in the development of SVERIGE, a micro simulation model that handles the full population of Sweden on an individual level. For the development of SVERIGE, a unique database with detailed individual biographies of all individuals in the population for a period of forty years was used (Holm *et al.*). In SVERIGE each modelled individual can carry through a number of acts each year: relocate, have children, find or change work and income, get an education, find a partner etc. In comparison with other similar models, SVERIGE is at least ten times faster. Also unique is the geographical resolution: each individual is located with a precision of 100 meters. Many of the processes that are modelled – such as the education, work, accommodation and partner markets – are very scale dependant, so it is a great advantage to be able to run simulations on millions of people individually in the model.

This technology is now used in a project with the aim to develop a similar model, tentatively called KINA, for the whole of China. The project includes both Chinese and Swedish scientists, working in cooperation. Presently we are working with a one percent selection of the Chinese population on an individual level, this gives a sample with approximately 13 million individuals. The data originates from the

1 Umeå University, Sweden
2 The Chinese Academy of Agricultural Sciences
3 Institute of Natural Resources and Regional Planning, China

1982 and 1990 censuses and is complemented with detailed demographic information from the year 2000. This project aims to develop KINA to such a level, that scenarios of population and labour development in 2500 counties can be realised and questions such as those previously presented can be investigated. Adapting SVERIGE to the requirements of KINA will generalise it in such a way that this unique technology could in the future be used in yet other environments.

Model Design

The KINA model is based on the SVERIGE (System for Visualising Economic and Regional influences in Governing) model, developed at the Spatial Modelling Centre (SMC) at Department of Social and Economic Geography, Umeå University (Holm *et al.*). Basically it *is* the SVERIGE model, only adapted to Chinese conditions, so much of what follows is, at least in principle, applicable on both models. What differs the two models is some changes in the way families are handled (Chinese families are usually larger than Swedish, since they may include more generations) and that the data in KINA is more rudimentary, in some few cases data from SVERIGE was used since none from China was at hand. The individual data is rudimentary both in the number of individuals represented (only one per cent of the entire population) and in the number and quality of the attributes. The same is true for the event modules. It will now be a continuous task to refine the data and event modules, and to keep them up to date.

The basic agent in the model is the individual person, but the family (household) is the operational unit. Using the family as the basic unit is natural, since its members share many resources such as income, housing etc. Families are not static, however. They grow or shrink in size, they disappear and new ones are created, and so on. They can also move, either migrate within the model to a new geographic location with other resources and limitations or emigrate out of the model – and effectively disappear.

The demographical attributes are correlated with other demographical parameters and environmental factors. Right now, the model is bare bones, more precise parameter estimations will be required. Presently they are estimated by probability tables and rules. The probability tables mainly stem from the 1990 and 2000 censuses data. The rules are based on demographical theories or related studies.

The model is time driven and uses a time step of one year. This means that each individual is checked for any changes in his or her status once a "year" – the main part of the model is a loop that, family by family, checks each individual for a number of possible status changes and calculates the consequences of such changes. Possible status changes are determined in event modules: ageing, mortality, emigration, fertility, education, partnership, leaving home, divorce, migration, employment, and immigration.

Ageing

At the beginning of each simulation step, all time-related attributes are incremented by one year.

Mortality

This module terminate lives. The model is dynamic, so as an individual ages the death risk changes. The factors that presently determines the probability of death are just age, gender, and the individual's environment (rural or urban). We plan to soon let the province of the individual be another factor, since it is obvious that there are great variations in mortality between the provinces. The death probability is given by age in four tables: urban women, urban men, rural women, and rural men. The probabilities were calculated from data provided by the 1990 and 2000 censuses.

As the consequence of a death, the individual is of course removed. Further, the marital status of the spouse is changed and some family roles might change if a new head of family is required.

In the future, this module will be replaced by a binary logistic module where the attributes age, age squared, education level, gender, and province will be included in the equation.

Emigration

Emigration is treated as if the whole family died – it is removed from the model. The factors to determine the probability of emigration is the age and gender of the family head, the probability is given by age in two tables: one for women and one for men. Since the present yearly emigration from China is not more than %, this is not a very important module. However, education is probably also an important factor.

Fertility

This module and, to a much smaller extent the immigration module, populate the model with new individuals. Fertility is in the model influenced by gender (obviously), age, and the number of previous children. The latter is especially important in China since the number of children in a family is regulated since the 1970's. All families are allowed one child, in rural areas families are under certain restrictions allowed to have more children. Only women of age 18–46 have children and a fourth child is never created in the model. The source for the probability figures is the 1990 census and is given by age in three tables: zero, one, and two previous children.

Education

In the model the level of education is divided into eight levels. At each level (except the last one, of course) there is a certain age-dependant probability that the individual will continue to the next level. The probability is given by age in seven tables, corresponding to the first seven education levels.

Partnership

Creating new families is a three stage process:

1. First it is determined if an individual wants to find a partner. To be eligible the civil status must be either single, divorced, or widowed. The probability is the given by age in table(s).
2. Then couples are created. People are paired according to gender (obviously), age, education level, and location (pairing is only done locally, by county).
3. Lastly, the couple and their families are merged – or, if there was no partner, the individual is returned to the original household with no change of any attribute. A returned individual will try again at step one at the next iteration (year).

Leaving home

This module creates new families by letting children move from their parental household. The probability of moving out is gender-independent, so it's given by age in a single table.

Divorce

The module splits one family into two. The new family will consist of the man and will be located in the same area as the original family. The probability of divorce depends on the woman's age and is given by a table.

Migration

The relocation of a family is governed by this module. The entire family is either relocated or not, it is never split by this module. The probability of relocation depends of the family head's age and is give by two tables: one for women and one for men. The probability is based on data from the 1990 and 2000 national censuses.

Employment

In the model an individual belong to one of four categories: have never had a job, had a job but is presently without, have a job, or is retired. The probability of a transition between these categories are given by age in four tables: getting the first job, quitting, reemployment, and retirement.

Immigration

An immigration pool was created from which new immigrants are cloned. This module populate the model with individuals that already have a history. The number of people added is given by age in two tables: one for women and one for men.

Technical Aspects of the Model

The model runs on both Windows machines and Unix/Linux machines – and it gives reasonable predictions.

Since the number of agents is great, various techniques are used to minimize the amount of memory used by each agent. Some attributes, such as the geographical location or number of children, are shared by all members in a family. Some attributes make sense only to active adults, but not to children or pensioners. For example, the level of education is not relevant for either children or pensioners – although for different reasons. However, in general agents have the following attributes:

Attribute	Bits	Children	Old people	Comment
Family identifier	27	x	x	saved by family
Region	12	x	x	saved by family
Personal identifier	25	x	x	
Years in school	4	x	x	
Birth year	9	x	x	
Rural or urban	1	x	x	
In education	1	x	x	
Alive	1	x	x	
Gender	1	x	x	
Relation in family	3	x	x	head, parent, child, etc.
Ethnic group	6	x	x	
Marital status	3		x	
Educational level	3		x	
Educational sector	3		x	
Out of labour	1			
Working	1			
Unemployed	1			

One per cent sample of the 1990 census

The following data or data sets are used in the China population modelling:

1. the national demographical census data in 1990,
2. aggregation data of 2000 national demographical census,

3. county-level demographical database,
4. provincial migration database,
5. county-level agro-eco-environmental database, and
6. map of county-level administrative regions.

The national demographical census data in 1990

The 1990 population Census of China was conducted by the State Statistics Bureau of China. This data set (1% sampling) was prepared taking villages as sampling unit, and contains a record for each household and supplies variables describing the location, type, and composition of the households. Each household record is followed by a record for each individual residing in the household. Information on individual includes demographic characteristics, occupation, literacy, ethnicity, and fertility. The record length for both household and individual observations are 52 bytes.

Originally, this data was organized for each province in ASCII format. We developed a program to integrate all the data files into one data file for whole China. All these data files were also concatenated to a file and then transported into SPSS software for further analysis and data extraction. Overall, there are 11,835,908 entities in the data set.

Aggregation data of the 2000 national demographical census

The 5:th National Demographic Census in China was committed on November 1, 2000. The aggregation data set was compiled by the Population Census Office of the State Council and National Bureau of Statistics of China, and it was published in 2002.

County-level demographical database

There are the frequently used demographic indices in 1990 for more than 2000 counties from 30 provinces, autonomous regions and municipalities. These indices include total population, sex ratio, size of family, population in cities, population in towns, the ratio of illiterate and semi-illiterate among the population over 15, the number of college-educated people per 10,000 people, the number of middle-school-educated people per 10,000 people, population aged between 0–14, population aged 15–49, population aged 50–65, population older than 65.

Provincial migration database

This database summarized the population migration data from 1949. It included the outputs from the 4:th national demographic census in 1990, and some other large-scale sampling survey since the 1980s, such as the migration survey of 74 cities in 1986, 1‰ national sampling of population in 1987, the national sampling survey on anti-pregnancy in 1988, the sampling on pregnancy and delivery in 1992, and other related data from police's registering system.

County-level agro-eco-environmental database

The database is from the Information Center, Ministry of Agriculture of China. It contains the statistics for agriculture production in 2523 counties in China. The data is for 1986–2001. The data items includes population, crop land acreage, acreage and yield for each crop, GDP, etc.

Estimation and validation

The model is compared to the actual population development in Figure **??**.

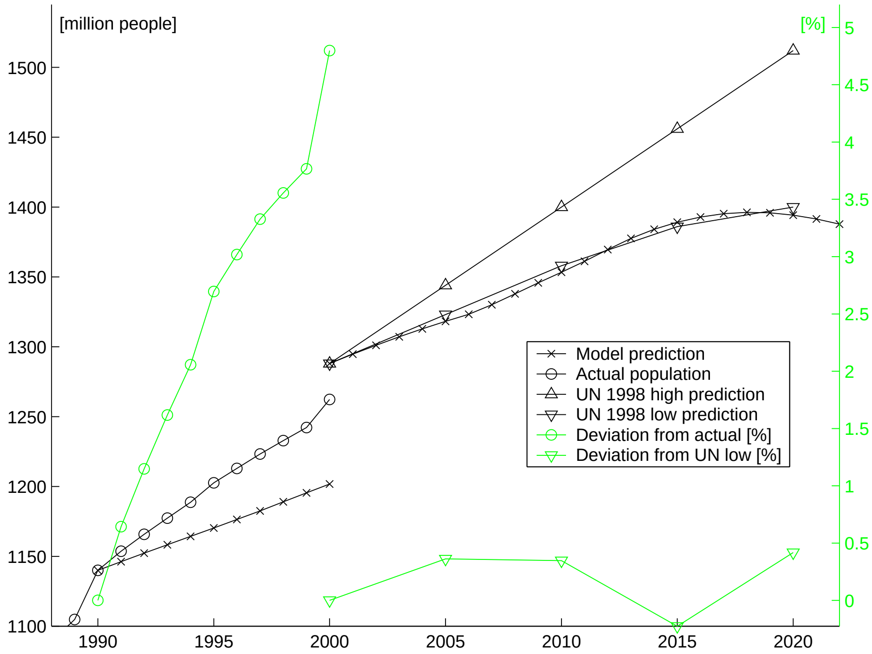

Figure 1: The diagram shows a comparison of the model with China's actual population figures and UN predictions. Since the model only uses a subset of the total population, the results have been normalized as to be comparable. The model's prediction for 1990–2000 was normalized against the actual population 1990. Likewise, the model's prediction for 2000–2020 was normalized against the UN 1998 prediction for 2000. (Source: *XXXXXX.*)

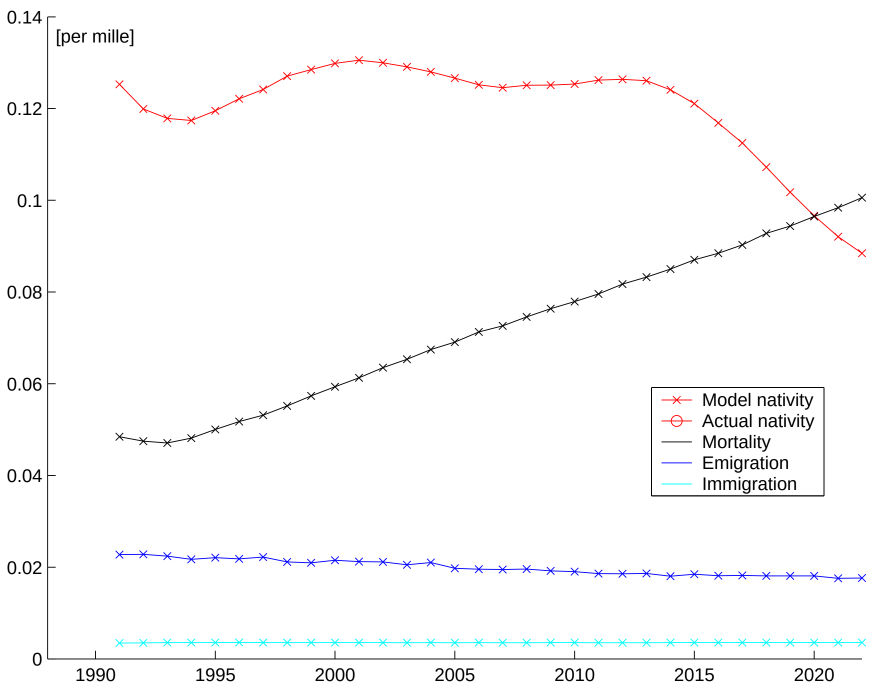

Figure 2: The diagram shows a comparison of the model with actual figures with respect to nativity, immigration, mortality, and emmigration. (Source: *XXXXXX.*)

Discussion

Method

Applications

Bibliography

Holm Holm, E., Holme, K., Mäkilä, K., Mattsson-Kauppi, M., Mörtvik, G. *The* SVERIGE *spatial microsimulation model – Content, validation, and example applications*. Kulturgeografiska institutionen/SMC.

Appendix: Tables

Total Population

Year	Actual values comparison			UN 1998 prediction comparison		
	Model [million]	Actual [million]	Deviation [%]	Model [million]	UN [million]	Deviation [%]
1990	1140.00	1140.00	0.00	1221.86	–	–
1991	1146.20	1153.62	0.64	1228.51	–	–
1992	1152.31	1165.70	1.15	1235.06	–	–
1993	1158.31	1177.36	1.62	1241.49	–	–
1994	1164.26	1188.71	2.06	1247.86	–	–
1995	1170.22	1202.63	2.70	1254.25	–	–
1996	1176.34	1212.95	3.02	1260.81	–	–
1997	1182.53	1223.22	3.33	1267.44	–	–
1998	1188.98	1232.82	3.56	1274.36	–	–
1999	1195.40	1242.19	3.77	1281.24	–	–
2000	1201.71	1262.28	4.80	1288.00	1288.00	0.00
2001	1207.93	–	–	1294.67	–	–
2002	1213.85	–	–	1301.02	–	–
2003	1219.55	–	–	1307.13	–	–
2004	1224.82	–	–	1312.77	–	–
2005	1229.89	–	–	1318.21	1323.00	0.36
2006	1234.56	–	–	1323.21	–	–
2007	1240.98	–	–	1330.10	–	–
2008	1248.20	–	–	1337.83	–	–
2009	1255.61	–	–	1345.77	–	–
2010	1262.62	–	–	1353.29	1358.00	0.35
2011	1269.88	–	–	1361.07	–	–
2012	1277.74	–	–	1369.49	–	–
2013	1285.26	–	–	1377.55	–	–
2014	1291.38	–	–	1384.11	–	–
2015	1296.05	–	–	1389.12	1386.00	−0.22
2016	1299.51	–	–	1392.83	–	–
2017	1301.75	–	–	1395.23	–	–
2018	1302.63	–	–	1396.17	–	–
2019	1302.28	–	–	1395.80	–	–
2020	1300.76	–	–	1394.16	1400.00	0.42

Nativity, Immigration, Mortality, and Emigration

Year	Nativity [‰]		Immigration [‰]		Mortality [‰]		Emigration [‰]	
	Model	Actual	Model	Actual	Model	Actual	Model	Actual
1990	–	–	–	–	–	–	–	–
1991	0.125	–	0.003	–	0.048	–	0.023	–
1992	0.120	–	0.003	–	0.047	–	0.023	–
1993	0.118	–	0.004	–	0.047	–	0.022	–
1994	0.117	–	0.004	–	0.048	–	0.022	–
1995	0.120	–	0.004	–	0.050	–	0.022	–
1996	0.122	–	0.004	–	0.052	–	0.022	–
1997	0.124	–	0.004	–	0.053	–	0.022	–
1998	0.127	–	0.004	–	0.055	–	0.021	–
1999	0.129	–	0.004	–	0.057	–	0.021	–
2000	0.130	–	0.004	–	0.059	–	0.021	–
2001	0.131	–	0.004	–	0.061	–	0.021	–
2002	0.130	–	0.004	–	0.064	–	0.021	–
2003	0.129	–	0.004	–	0.065	–	0.021	–
2004	0.128	–	0.004	–	0.067	–	0.021	–
2005	0.127	–	0.004	–	0.069	–	0.020	–
2006	0.125	–	0.004	–	0.071	–	0.020	–
2007	0.125	–	0.004	–	0.073	–	0.019	–
2008	0.125	–	0.004	–	0.075	–	0.020	–
2009	0.125	–	0.004	–	0.076	–	0.019	–
2010	0.125	–	0.004	–	0.078	–	0.019	–
2011	0.126	–	0.004	–	0.080	–	0.019	–
2012	0.126	–	0.004	–	0.082	–	0.019	–
2013	0.126	–	0.004	–	0.083	–	0.019	–
2014	0.124	–	0.004	–	0.085	–	0.018	–
2015	0.121	–	0.004	–	0.087	–	0.018	–
2016	0.117	–	0.004	–	0.088	–	0.018	–
2017	0.112	–	0.004	–	0.090	–	0.018	–
2018	0.107	–	0.004	–	0.093	–	0.018	–
2019	0.102	–	0.004	–	0.094	–	0.018	–
2020	0.097	–	0.004	–	0.096	–	0.018	–
2021	0.092	–	0.004	–	0.098	–	0.018	–
2022	0.088	–	0.004	–	0.101	–	0.018	–