文章编号：2096-1618(2020)02-0239-05

Logistic 模型的改进与中国人口预测

陈　霞，　肖　岚
(成都信息工程大学应用数学学院，四川 成都 610225)

摘要：在 Logistic 种群增长模型中引入适当形式的收获函数，得到改进模型并求出精确解，利用最小二乘法的原理和 Matlab 软件对改进模型进行参数估计，通过曲线拟合对比验证了该参数估计的方法能取得较好的效果，并将其用于中国人口预测。与其他文献预测结果相比较，文中预测结果更加吻合中国人口真实数据。

关　键　词：Logistic 模型；收获函数；最小二乘法；中国人口预测

中图分类号：O212　　　　**文献标志码**：A

doi：10.16836/j.cnki.jcuit.2020.02.016

0　引言

为了研究人口及物种(鱼塘中的鱼群、深林中的植物等)数量的变化规律，提出了许多数学模型，常见的是指数增长模型和阻滞增长模型(Logistic 增长模型)。

指数增长模型[1-3]：

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=rx(t)\quad r>0$$

Logistic 增长模型[1-3]：

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=rx(t)\left(1-\frac{x(t)}{C}\right)\quad r>0,C>0$$

其中，$x(t)$ 表示 t 时刻种群的数量，r 表示种群的固有增长率(即出生率减去死亡率)，C 表示环境承载力(即环境所能容纳的最大种群数量)。

为了提高 Logistic 增长模型在预测种群数量方面的准确性与实用性，由实际需求出发对 Logistic 增长模型进行改进，推广到具有收获函数的形式。Laham 等[4]在 Logistic 增长模型中引入收获函数，将经典的 Logistic 增长模型改为：

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=rx(t)\left(1-\frac{x(t)}{C}\right)-H(t)\quad r>0,C>0$$

其中 $H(t)$ 表示收获函数。将该模型用于制定鱼的捕获策略中，当收获函数 $H(t)$ 为常值函数时，说明在某一时刻捕获固定数量的鱼类；当收获函数 $H(t)$ 为周期函数时，保证任意时刻鱼类都不会濒于灭绝。Alfred[5]研究了濑鱼养殖的捕获策略，讨论了收获函数是常值函数时的收获增长模型：

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=rx(t)\left(1-\frac{x(t)}{C}\right)-h$$

该模型表示在给定时间间隔内捕获一定数量的濑鱼，估计了保护濑鱼免于灭绝的最佳捕获量，为制定渔业捕捞策略提供了依据。

1　模型的改进及求解

随着研究的深入，一些学者认为种群的增长率也会随着自然环境的改变而变化，为了更好地体现种群在复杂环境中的变化状况，需要将增长率 r 表示为一个关于时间 t 的函数 $r(t)$，且收获函数 $H(t)$ 应与增长率 $r(t)$ 相关。假设收获函数 $H(t)$ 与增长率 $r(t)$ 存在一定的线性关系，即存在一个常数 k 使得 $H(t)=kr(t)$，于是得到一个改进的 Logistic 增长模型：

$$\begin{cases}\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=r(t)x(t)\left(1-\dfrac{x(t)}{C}\right)-kr(t)\\ x(0)=x_0\end{cases}\tag{1}$$

针对上述改进模型，可通过如下方式求模型的解[4]。令 $x(t)=y(t)+\delta$，则模型(1)可转变为

$$\begin{cases}\dfrac{\mathrm{d}y(t)}{\mathrm{d}t}=r(t)(y(t)+\delta)\left(1-\dfrac{y(t)+\delta}{C}\right)-kr(t)\\ y_0=x_0-\delta\end{cases}\tag{2}$$

即

$$\begin{cases}\dfrac{\mathrm{d}y(t)}{\mathrm{d}t}=r(t)y(t)\left(1-\dfrac{y(t)+2\delta}{C}\right)+r(t)\left(\delta-\dfrac{\delta^2}{C}-k\right)\\ y_0=x_0-\delta\end{cases}\tag{3}$$

令 $r(t)\left(\delta-\dfrac{\delta^2}{C}-k\right)=0$，又 $r(t)\neq0$，则

$$\delta=\frac{C}{2}\pm\sqrt{\frac{C^2}{4}-kC}$$

变成求解以下方程：

$$\begin{cases}\dfrac{\mathrm{d}y(t)}{\mathrm{d}t}=r(t)y(t)\left(1-\dfrac{y(t)+2\delta}{C}\right)\\ y_0=x_0-\delta\end{cases}\tag{4}$$

对(4)式分离变量并整理可得

收稿日期：2019-07-10
基金项目：四川省科技计划资助项目(2018JY0169)

$$y=\frac{y_0(C-2\delta)\mathrm{e}^{\frac{(C-2\delta)}{C}\int_0^t r(s)\mathrm{d}s}}{C-2\delta-y_0+y_0\mathrm{e}^{\frac{(C-2\delta)}{C}\int_0^t r(s)\mathrm{d}s}} \tag{5}$$

于是模型(1)具有精确解:

$$x(t)=\frac{Cx_0-2\delta x_0-C\delta+2\delta^2}{x_0-\delta+(C-\delta-x_0)e^{\frac{2\delta-C}{C}\int_0^t r(s)\mathrm{d}s}}+\delta \tag{6}$$

其中 $\delta=\frac{C}{2}\pm\sqrt{\frac{C^2}{4}-kC}$。

当收获函数 $H(t)$ 为非零常值函数时,此时 $r(t)$ 也是一个非零常值函数,于是模型(1)精确解可表示为

$$x(t)=\frac{Cx_0-2\delta x_0-C\delta+2\delta^2}{x_0-\delta+(C-\delta-x_0)\mathrm{e}^{-At}}+\delta \tag{7}$$

其中,$A=\frac{r(C-2\delta)}{C}$,$\delta=\frac{C}{2}\pm\sqrt{\frac{C^2}{4}-kC}$。

2 模型的参数估计与曲线拟合

2.1 模型的参数估计

模型中参数估计的准确性直接影响人口预测的准确性,考虑当增长率 $r(t)$ 和收获函数 $H(t)$ 都是非零常值函数时求得的解,对它进行参数估计。采用最小二乘法的原理,利用 Matlab 中 lsqcurvefit 函数对参数进行估计与反复修正,建立相对合理的预测模型[7-8]。最小二乘法的原理就是让观测值与实际值之间的残差平方和最小,残差平方和越小,说明模型预测值与实际值之间差异越小,模型更实用。

如对于模型

$$x(t)=\frac{Cx_0-2\delta x_0-C\delta+2\delta^2}{x_0-\delta+(C-\delta-x_0)\mathrm{e}^{-At}}+\delta$$

对式(7)中 3 个未知参数 C,δ,A 进行估计,首先求出观测值与实际值之间的最小残差平方和:

$$Q=\min\sum_{t=1}^{n}u_t^2=\min\sum_{t=1}^{n}(x_t-x(t))^2 \tag{8}$$

其中,x_t 表示 t 时刻的实际人口数量,$x(t)$ 表示 t 时刻的预测人口数量。

再用式(8)分别对 C,δ,A 求导等于零,建立方程组:

$$\begin{cases}\frac{\partial Q}{\partial C}=0\\ \frac{\partial Q}{\partial \delta}=0\\ \frac{\partial Q}{\partial A}=0\end{cases} \tag{9}$$

从而求出 C,δ,A 的估计值。但在求解过程中发现式(9)的计算过于烦琐较有难度,于是考虑借助数学软件,调用 Matlab 中的函数 lsqcurvefit 实现最小二乘非线性拟合,通过迭代求解模型的待定参数,从而使残差平方和最小,进而完成最优参数估计。

根据国家统计局官网统计数据,取中国 1950–2018 年的人口数据如表 1 所示。

表 1 中国人口数据(1950–2018 年)

时间 t	年份	实际总人口数/亿	时间 t	年份	实际总人口数/亿	时间 t	年份	实际总人口数/亿
$t=0$	1950	5.5196	$t=23$	1973	8.8761	$t=46$	1996	12.2389
$t=1$	1951	5.6300	$t=24$	1974	9.0409	$t=47$	1997	12.3626
$t=2$	1952	5.7482	$t=25$	1975	9.1970	$t=48$	1998	12.4761
$t=3$	1953	5.8796	$t=26$	1976	9.3267	$t=49$	1999	12.5786
$t=4$	1954	6.0266	$t=27$	1977	9.4774	$t=50$	2000	12.6743
$t=5$	1955	6.1465	$t=28$	1978	9.6259	$t=51$	2001	12.7627
$t=6$	1956	6.2780	$t=29$	1979	9.7542	$t=52$	2002	12.8453
$t=7$	1957	6.4238	$t=30$	1980	9.8705	$t=53$	2003	12.9227
$t=8$	1958	6.5346	$t=31$	1981	10.0072	$t=54$	2004	12.9988
$t=9$	1959	6.6012	$t=32$	1982	10.1654	$t=55$	2005	13.0756
$t=10$	1960	6.6207	$t=33$	1983	10.3008	$t=56$	2006	13.1448
$t=11$	1961	6.6457	$t=34$	1984	10.4357	$t=57$	2007	13.2129
$t=12$	1962	6.7295	$t=35$	1985	10.5851	$t=58$	2008	13.2802
$t=13$	1963	6.9172	$t=36$	1986	10.7507	$t=59$	2009	13.3450
$t=14$	1964	7.0499	$t=37$	1987	10.9300	$t=60$	2010	13.4091
$t=15$	1965	7.2538	$t=38$	1988	11.1026	$t=61$	2011	13.4735
$t=16$	1966	7.4206	$t=39$	1989	11.2704	$t=62$	2012	13.5404
$t=17$	1967	7.6032	$t=40$	1990	11.4333	$t=63$	2013	13.6072
$t=18$	1968	7.8198	$t=41$	1991	11.5823	$t=64$	2014	13.6782
$t=19$	1969	8.0335	$t=42$	1992	11.7171	$t=65$	2015	13.7462
$t=20$	1970	8.2542	$t=43$	1993	11.8517	$t=66$	2016	13.8271
$t=21$	1971	8.4779	$t=44$	1994	11.9850	$t=67$	2017	13.9008
$t=22$	1972	8.6727	$t=45$	1995	12.1121	$t=68$	2018	13.9538

模型(7)中含有3个未知参数 C,δ,A,一般可以将3组已知数据代入式(7)建立方程组得到未知参数值,但这样的值并不是最优的,于是可将这样求得的参数值作为参数估计的初始值,再利用 Matlab 软件中的函数 lsqcurvefit 得到最优的参数估计值。

具体求解过程如下:

首先,从表1中取数据代入式(7),利用 Matlab 软件求解未知参数。令 $x_0=5.5196$,将 $x(20)=8.2542$,$x(40)=11.4333$,$x(60)=13.4091$ 代入式(7)可得 $C=18.0571$,$\delta=3.4708$,$A=0.0603$,则式(7)表示为

$$x_1(t)=\frac{22.7734364}{2.0488+9.0667e^{-0.0603t}}+3.4708 \qquad (10)$$

在此基础上,调用 Matlab 软件中的函数 lsqcurvefit 采用非线性最小二乘对式(7)3个参数 C,δ,A 作最优估计。这里为了方便表示,对式(7)中的参数进行处理,令 $y=x(t)$,$a(1)=C$,$a(2)=\delta$,$a(3)=A$,将 $x_0=5.5196$代入式(7)可得

$$y=\frac{[a(1)-2a(2)][5.5196-a(2)]}{5.5196-a(2)+[a(1)-a(2)-5.5196]e^{-a(3)t}}+a(2) \qquad (11)$$

将 $C=18.0571$,$\delta=3.4708$,$A=0.0603$作为参数估计的初值,目的就是利用最小二乘法寻求更加理想的参数 $a(1)$,$a(2)$,$a(3)$,使模型更加优化。

利用 Matlab 求解程序如下:

函数文件 logistic:

```
function [ y ] = logistic2( a,t )
y=(a(1)-2 * a(2)) * (5.5196-a(2))/(5.5196-a(2)+(a(1)-a(2)-5.5196) * exp(-a(3) * t))+a(2)
end
```

主函数:

```
>>t=[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68];
y=[5.5196 5.6300 5.7482 5.8796 6.0266 6.1465 6.2780 6.4238 6.5346 6.6012 6.6207 6.6457 6.7295 6.9172 7.0499 7.2538 7.4206 7.6032 7.8198 8.0335 8.2542 8.4779 8.6727 8.8761 9.0409 9.1970 9.3267 9.4774 9.6259 9.7542 9.8705 10.0072 10.1654 10.3008 10.4357 10.5851 10.7507 10.9300 11.1026 11.2704 11.4333 11.5823 11.7171 11.8517 11.9850 12.1121 12.2389 12.3626 12.4761 12.5786 12.6743 12.7627 12.8453 12.9227 12.9988 13.0756 13.1448 13.2129 13.2802 13.3450 13.4091 13.4735 13.5404 13.6072 13.6782 13.7462 13.8271 13.9008 13.9538];% 人口数量
a0=[18.0571,3.4708,0.0603];% 参数初值
[a,resnorm]=lsqcurvefit(@ logisti,a0,t,y)
```

此时求得的参数最优估计值为:$a(1)=18.2746$,$a(2)=3.6727$,$a(3)=0.0623$及在 a0 处的目标最小二乘表达式(即所谓的残差)为 resnorm=0.5192。再将得到的 $a(1)$,$a(2)$,$a(3)$作为参数初值再次代入上述程序中,进行二次迭代,此时求得的参数最优估计值为:$a(1)=18.2746$,$a(2)=3.6727$,$a(3)=0.0623$,以及在 a0 处的目标最小二乘表达式(即所谓的残差)为 resnorm=0.5192。发现第二次代入初值计算,残差结果与第一次一样,停止迭代。经过计算,当 $a(1)=18.2746$,$a(2)=3.6727$,$a(3)=0.0623$ 时,残差取得最小,所以将最优估计结果代入式(7)可得

$$x_2(t)=\frac{20.18513948}{1.8469+9.0823e^{-0.0623t}}+3.6727 \qquad (12)$$

2.2　曲线拟合比较

为了检验预测结果的合理性,并说明利用最小二乘法做参数估计的优越性,分别将表1中1950-2018年中国总人口数量的实际数据与求得的式(10)和式(12)中 $x(t)$的函数图形进行拟合比较。

利用 Matlab 求解程序如下:

(i)1950-2018 年中国人口数量的实际数据与式2704(10)中 $x_1(t)$的函数图形进行拟合:

```
t1=1950:2018;
X1=[5.5196 5.6300 5.7482 5.8796 6.0266 6.1465 6.2780 6.4238 6.5346 6.6012 6.6207 6.6457 6.7295 6.9172 7.0499 7.2538 7.4206 7.6032 7.8198 8.0335 8.2542 8.4779 8.6727 8.8761 9.0409 9.1970 9.3267 9.4774 9.6259 9.7542 9.8705 10.0072 10.1654 10.3008 10.4357 10.5851 10.7507 10.9300 11.1026 11.2704 11.4333 11.5823 11.7171 11.8517 11.9850 12.1121 12.2389 12.3626 12.4761 12.5786 12.6743 12.7627 12.8453 12.9227 12.9988 13.0756 13.1448 13.2129 13.2802 13.3450 13.4091 13.4735 13.5404 13.6072 13.6782 13.7462 13.8271 13.9008 13.9538];t=1950:2050;
X=22.7734364/(2.0488+9.0667 * exp(-0.0603 * (t-1950)))+3.4708;
plot(t1,x1,´-´,t,x,´.´)
```

通过运行该程序,结果如图1所示。

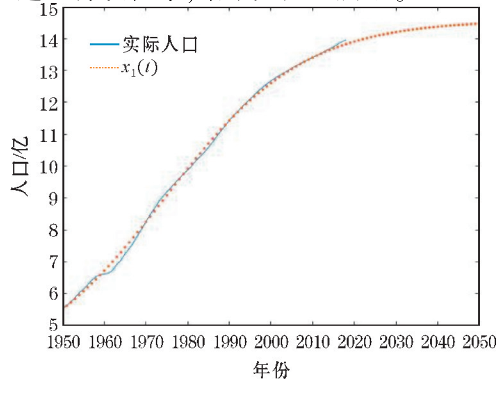

图1　实际总人口数据与 $x_1(t)$ 的函数图形比较

(ⅱ)1950-2018 年中国人口数量的实际数据与式(12)中 $x_2(t)$ 的函数图形进行拟合:

```
t1=1950:2018;
x1 = [5.5196 5.6300 5.7482 5.8796 6.0266 6.1465 6.2780 6.4238 6.5346 6.6012 6.6207 6.6457 6.7295 6.9172 7.0499 7.2538 7.4206 7.6032 7.8198 8.0335 8.2542 8.4779 8.6727 8.8761 9.0409 9.1970 9.3267 9.4774 9.6259 9.7542 9.8705 10.0072 10.1654 10.3008 10.4357 10.5851 10.7507 10.9300 11.1026 11.2704 11.4333 11.5823 11.7171 11.8517 11.9850 12.1121 12.2389 12.3626 12.4761 12.5786 12.6743 12.7627 12.8453 12.9227 12.9988 13.0756 13.1448 13.2129 13.2802 13.3450 13.4091 13.4735 13.5404 13.6072 13.6782 13.7462 13.8271 13.9008 13.9538];
t=1950:2050;
x=20.18513948/(1.8469+9.0823*exp(-0.0623*(t-1950)))+3.6727;
plot(t1,x1,'-',t,x,'.')
```

通过运行该程序,结果如图 2 所示。

通过观察,比较图 1 与图 2,虽然 $x_1(t)$ 和 $x_2(t)$ 的函数图形都与实际人口数据的走向大致吻合,但显然图 2 与实际的人口数据拟合程度更高更准确,说明利用最小二乘法做参数估计使参数求解更加理想。

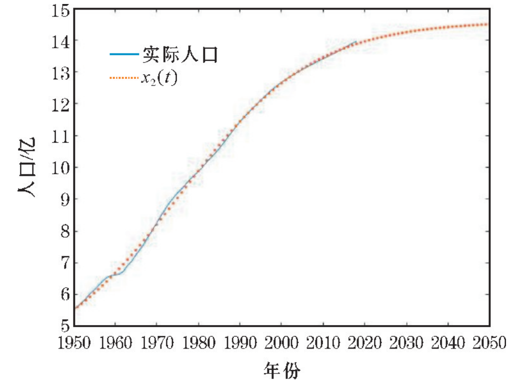

图 2 实际总人口数据与 $x_2(t)$ 的函数图形比较

3 将改进的模型用于中国人口预测

由前文可知,改进的 Logistic 增长模型

$$x_2(t)=\frac{20.18513948}{1.8469+9.0823e^{-0.0623t}}+3.6727$$

能很好地与中国人口数据相拟合,当 t 取不同值时,可以得到不同时间下人口数量的预测结果,所以可以用于中国人口预测。

为了证明改进的 Logistic 模型以及利用最小二乘进行参数估计的方法用于人口预测的效果较好,将 1994-2014 年的中国人口数据与文献[9-10]进行对比,对比预测结果见表 2。

表 2 1994-2014 年中国人口预测对比

年份	实际人口/亿	文中预测值	绝对误差	文献[9]预测值	绝对误差	文献[10]预测值	绝对误差
1994	11.9850	11.9703	0.0147	11.3818	0.6032	11.8755	0.1095
1995	12.1121	12.0928	0.0193	11.5137	0.5984	11.9835	0.1286
1996	12.2389	12.2112	0.0277	11.6448	0.5941	12.0908	0.1481
1997	12.3626	12.3255	0.0371	11.7748	0.5878	12.1974	0.1652
1998	12.4761	12.4357	0.0404	11.9039	0.5722	12.3033	0.1728
1999	12.5786	12.5419	0.0367	12.0318	0.5468	12.4085	0.1701
2000	12.6743	12.6440	0.0303	12.1587	0.5156	12.5129	0.1614
2001	12.7627	12.7422	0.0205	12.2844	0.4783	12.6166	0.1461
2002	12.8453	12.8363	0.0090	12.4089	0.4364	12.7195	0.1258
2003	12.9227	12.9266	0.0039	12.5322	0.3905	12.8216	0.1011
2004	12.9988	13.0131	0.0143	12.6542	0.3446	12.9229	0.0759
2005	13.0756	13.0959	0.0203	12.7748	0.3008	13.0233	0.0523
2006	13.1448	13.1749	0.0301	12.8942	0.2506	13.1230	0.0218
2007	13.2129	13.2505	0.0376	13.0122	0.2007	13.2218	0.0089
2008	13.2802	13.3225	0.0423	13.1287	0.1515	13.3197	0.0395
2009	13.3450	13.3913	0.0463	13.2439	0.1011	13.4168	0.0718
2010	13.4091	13.4567	0.0476	13.3575	0.0516	13.5130	0.1039
2011	13.4735	13.5190	0.0455	13.4697	0.0038	13.6083	0.1348
2012	13.5404	13.5783	0.0379	13.5804	0.0400	13.7027	0.1623
2013	13.6072	13.6347	0.0275	13.6896	0.0824	13.7963	0.1891
2014	13.6782	13.6882	0.0100	13.7972	0.1190	13.8889	0.2107

本文平均误差 0.0285238095238095,文献[9]平均误差 0.33187619047619,文献[10]平均误差,0.119033333333333

通过表 2 将文中与其他文献中国人口预测值进行比较,发现本文、文献[9]、文献[10]的人口预测值与实际值之间的平均误差分别为:0.0285,0.3319,0.1190,明显本文改进的 Logistic 人口预测模型预测效果较好,更能反映中国人口的增长变化趋势。

该对比结果说明改进的 Logistic 模型在人口预测方面能更好地吻合中国人口真实数据,将它用于预测未来中国人口数量具有一定的合理性。于是将改进的 Logistic 模型用于未来中国人口预测,结果见表 3。

表 3　未来中国人口预测

年份	预测值/亿	年份	预测值/亿	年份	预测值/亿
2019	13.9174	2024	14.0921	2029	14.2238
2020	13.9563	2025	14.1215	2030	14.2459
2021	13.9932	2026	14.1493	2040	14.4080
2022	14.0280	2027	14.1756	2050	14.4971
2023	14.0609	2028	14.2004	…	…

4　结束语

在 Logistic 增长模型中引入收获函数,从而建立了一个新的改进模型,并利用最小二乘法原理对模型中相关未知参数进行估计,通过曲线拟合与数据比对,将改进模型用于中国人口预测,取得了较好效果,为未来中国人口预测提供一定的参考价值。

参考文献:

[1] 陈兰荪. 数学生态学模型与研究方法[M]. 北京:科学出版社,1985.

[2] 姜启源. 数学模型[M]. 北京:高等教育出版社,1993.

[3] 林振山. 种群动力学[M]. 北京:科学出版社,2006.

[4] M F Laham, I S Krishnarajah, J M Shariff. Fish Harvesting Management Strategies Using Logistic Growth Model[J]. Sains Malaysiana, 2012, 41(2):171-177.

[5] Alfred DACI, Fish Harvesting Models And Their Applications in a reservoir in Sanrada Albania[J]. JMEST,2016,3:5279-5282.

[6] 王高雄,周之铭. 常微分方程[M]. 北京:高等教育出版社,1983.

[7] 周品. MATLAB 数值分析应用教程[M]. 北京:电子工业出版社,2014.

[8] 郑小洋,罗其莉,姚蕾,等. 人口阻滞增长模型的参数估计与程序实现[J]. 数学学习与研究,2015(1):132.

[9] 李利利,曾晔,龙元芳,等. 中国人口 Logistic 预测模型[J]. 高师理科学刊,2014,34(6):19-22.

[10] 刘锋,何卓,谭祥勇. Richards 模型与 Logistic 模型在人口预测中的比较[J]. 重庆工商大学学报,2017,34(1):6-9.

Improved Logistic Model and China Population Forecast

CHEN Xia,　XIAO Lan

(College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Logistic model is improved by introducing appropriate harvesting function and the exact solution is obtained. The parameters involved in the model are estimated by the principle of least squares method and Matlab software. Through the comparison of curve fitting, it is verified that the method of parameter estimation can achieve better results, and it is applied to the population prediction of China. Compared with other prediction results, the results of this paper are more consistent with the real data of Chinese population, and have good practical significance.

Keywords: Logistic model; harvest function; least squares method; China population forecast