东北玉米热量指数预测方法研究(Ⅲ)*

——GM(1, 1)预测方法

郭建平[1], 陈玥熤[1,2], 庄立伟[3]

(1. 中国气象科学研究院, 北京 100081; 2. 广东省气象局, 广东 广州 510080; 3. 国家气象中心, 北京 100081)

摘　要: GM(1, 1)是开展时间序列环境要素变化趋势预测的有效方法之一。通过对东北地区玉米热量指数的分析, 建立了热量指数时间变化趋势的 GM(1, 1)预测模型, 各模型的平均预测精度虽低于逐步回归统计模型, 但也都达 91% 以上, 可以应用该模型的预测结果指导农业生产。

关键词: 东北地区; 玉米; 热量指数; 预测; GM(1, 1)模型

中图分类号: S426　　**文献标识码:** A　　**文章编号:** 1000 - 811X(2010)01 - 0023 - 04

0　引言

灰色模型(Grey Model), 简称 GM 模型, 是灰色系统理论的基本模型, 也是灰色系统理论的核心, 是灰色预测、决策、控制的基础。灰色预测, 是指根据过去及现在已知的或非确知的信息, 建立一个从过去引伸到未来的 GM 模型, 从而确定系统在未来发展变化的趋势, 并为规划决策提供依据[1]。灰色系统理论建模主要是找出系统内因素本身或因素之间的数学关系, 从而了解系统的动态行为和发展趋势。它是以灰色模块(是时间序列在时间数据平面上的连续曲线或逼近曲线与时间轴所围成的区域)为基础, 以微分拟合法而建成的模型。GM(n, h)表示对 h 个变量用 n 阶微分方程建立的模型。在 GM(n, h)模型中, 当 h≥2 时, 所建的 GM 模型不能做预测用, 只能用于分析因子之间的相互关系。做预测用的一般为 GM(n, 1)模型, 其中最重要的同时也是在实际中应用得最多的是一阶一元灰色模型, 即 GM(1, 1)模型。GM(1, 1)预测模型在各个领域都得到了广泛的应用[2-13], 取得了显著的效果。

1　建立 GM(1, 1)模型的原理

1.1　数据处理

灰色系统在建模时, 首先要对原始数据进行处理, 在一定程度上相对增强原始数据序列的确定性和相对减弱不确定性。根据文献[1]中的定理, 非负的原始时间序列的数据幅值变化无规律, 而累加生成后的序列不但非负, 并且单调递增, 即数据幅值变化有一定的规律。生成序列与原始序列相比, 确定性增强了。于是, 灰色系统把问题放在生成层次上求解, 建立的不是原始数据模型, 而是生成序列的数据模型。通过关联分析, 生成时序所接近或相像的函数, 就是寻找的生成函数, 据此建立被研究对象的模型。最后通过建立生成序列模型得到的预测值, 必须做还原处理。

使原随机序列变成有规律序列, 弱化其随机性, 并且为建模提供中间信息, 这种生成变换也就是灰色量的"白化", 累加或累减是生成运算的基本手段[1,14-16]。在 GM(1, 1)模型中, 一般只对数列做一次累加(1—AGO; Accumulated Generating Operation), 即对原始数列中各时刻的数据依次累加(原始数据要求均为非负数, 否则累加时会正负抵消, 达不到使数据序列随时间递增的目的)。设

* 收稿日期: 2009 - 05 - 31

基金项目: "十一五"国家科技支撑计划课题(2006BAD04B02)

作者简介: 郭建平(1963 -), 男, 江苏昆山人, 博士, 研究员, 主要从事农业气象灾害、气候变化影响等研究.

E-mail: gjp@ cams. cma. gov. cn

原始数列为：

$$X^{(0)} = \{ X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(n) \}。 \quad (1)$$

在本研究中 n 为年，一次累加后得：

$$X^{(1)}(t) = \sum_{k=1}^{t} X^{(0)}(k), t = 1,2,\cdots,n 。 \quad (2)$$

1.2 建立 GM(1，1)模型

累加后的生成序列 $X^{(1)}$ 满足微分方程：

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u。 \quad (3)$$

式(3)称为 GM(1，1)模型。其中 a 和 u 为待定系数，a 是模型的发展系数，u 为内生控制参数。此一阶单变量常微分方程的解为：

$$X^{(1)}(t+1) = \left(X^{(1)}(1) - \frac{u}{a}\right)e^{-at} + \frac{u}{a} 。 \quad (4)$$

对式(3)的求解过程，实际上是将微分方程差分化，再用最小二乘法求解。具体方法如下：

$X^{(1)}(1)$留做初值用，且 $X^{(1)}(1) = X^{(0)}(1)$，将 $X^{(1)}(2)$，$X^{(1)}(3)$，$\cdots X^{(1)}(n)$分别代入(3)式，并用差分形式写出，取等时间间隔 $\Delta t = 1$，得：

$$\frac{\Delta X^{(1)}(2)}{\Delta t} = \Delta X^{(1)}(2) = X^{(1)}(2) - X^{(1)}(1) = X^{(0)}(2) 。 \quad (5)$$

类似的：

$$\frac{\Delta X^{(1)}(3)}{\Delta t} = X^{(0)}(3), \cdots, \frac{\Delta X^{(1)}(n)}{\Delta t} = X^{(0)}(n) 。 \quad (6)$$

于是，式(3)改写为

$$\begin{aligned} &X^{(0)}(2) + aX^{(1)}(2) = u , \\ &X^{(0)}(3) + aX^{(1)}(3) = u , \\ &\cdots\cdots \\ &X^{(0)}(n) + aX^{(1)}(n) = u 。 \end{aligned} \quad (7)$$

把 $aX^{(1)}(t)$项移到右边，并写成向量的形式，得

$$\begin{aligned} &X^{(0)}(2) = (-X^{(1)}(2), 1)\begin{pmatrix} a \\ u \end{pmatrix} , \\ &X^{(0)}(3) = (-X^{(1)}(3), 1)\begin{pmatrix} a \\ u \end{pmatrix} , \\ &\cdots\cdots \\ &X^{(0)}(n) = (-X^{(1)}(n), 1)\begin{pmatrix} a \\ u \end{pmatrix} 。 \end{aligned} \quad (8)$$

由于$\frac{\Delta X^{(1)}}{\Delta t}$涉及到累加序列 $X^{(1)}$ 的两个时刻的值，一般取前后两个时刻的等权滑动平均代替 $X^{(1)}(t)$，写为矩阵表达式后得到：

$$\begin{pmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \vdots \\ X^{(0)}(n) \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}(X^{(1)}(2) + X^{(1)}(1)) & 1 \\ -\frac{1}{2}(X^{(1)}(3) + X^{(1)}(2)) & 1 \\ \vdots & \\ -\frac{1}{2}(X^{(1)}(n) + X^{(1)}(n-1)) & 1 \end{pmatrix} \begin{pmatrix} a \\ u \end{pmatrix} 。 \quad (9)$$

令 $Y = (X^{(0)}(2), X^{(0)}(3), \cdots, X^{(0)}(n))^T$

$$B = \begin{pmatrix} -\frac{1}{2}(X^{(1)}(2) + X^{(1)}(1)) & 1 \\ -\frac{1}{2}(X^{(1)}(3) + X^{(1)}(2)) & 1 \\ \vdots & \\ -\frac{1}{2}(X^{(1)}(n) + X^{(1)}(n-1)) & 1 \end{pmatrix}, \quad U = \begin{pmatrix} a \\ u \end{pmatrix} 。$$

则式(9)的矩阵式可以写为

$$Y = BU 。$$

此式的最小二乘估计为：

$$\hat{U} = \begin{pmatrix} \hat{a} \\ \hat{u} \end{pmatrix} = (B^T B)^{-1} B^T Y 。 \quad (10)$$

把估计值 $\hat{a}$ 和 $\hat{u}$ 代入(4)式，得到时间响应方程：

$$\hat{X}^{(1)}(t+1) = (X^{(0)}(1) - \frac{\hat{u}}{\hat{a}})e^{-\hat{a}t} + \frac{\hat{u}}{\hat{a}} 。 \quad (11)$$

当 $t = 1, 2, \cdots, n-1$ 时，由式(11)计算的是拟合值；当 $t \geqslant n$ 时，为预报值。

2 GM(1，1)模型的精度检验

检验 GM(1，1)模型，一般采用关联度、残差大小、后验差等方式检验。在本研究中采用关联度检验。

因为关联度是反映两个事物在发展过程中的关联程度，而描述事物发展过程比较合理的指标是相对变化速率，如果两个事物在发展过程中的相对变化速率基本一致，就可以认为两者有较好的关联度。

定义函数 $X(t)$和 $Y(t)$的关联函数为：

$$\zeta(t) = \frac{1}{1 + \left| \frac{1}{X}\frac{dX}{dt} - \frac{1}{Y}\frac{dY}{dt} \right|} 。 \quad (12)$$

$\zeta(t) \in [0, 1]$，$X(t)$与 $Y(t)$在 t 时刻的相对变化速率越接近，$\zeta(t)$就越大。结合本研究，式(12)改用差分的形式，取 $\Delta t = 1$，则原始数列 $X^{(0)}(t)$和预测值 $\hat{X}^{(0)}(t)$的关联函数为：

$$\zeta(t)=\frac{1}{1+\left|\frac{X^{(0)}(t+1)-X^{(0)}(t)}{X^{(0)}(t)}-\frac{\hat{X}^{(0)}(t+1)-\hat{X}^{(0)}(t)}{\hat{X}^{(0)}(t)}\right|},$$

$t=1, 2, \cdots, n-1$。（13）

则根据关联度的定义：

$$r=\frac{1}{n-1}\sum_{t=1}^{n-1}\zeta(t)\ 。\qquad (14)$$

可以计算出 $X^{(0)}(t)$ 和 $\hat{X}^{(0)}(t)$ 的关联度。

由式(13)和式(14)可知，这样定义的关联函数反映了每一时刻两数列相对变化速率的一致程度，而关联度则反映了特定时段内两数列相对变化速率一致程度的平均状况。

3　结果分析

研究中利用 1961 －2000 年的资料建立模型，并进行模型回代检验，2001 －2005 年的资料用于模型的试报检验。

根据对 GM(1，1)模型的描述，实际应用过程中，模型的建立就是要求算模型中的 $\hat{a}$ 和 $\hat{u}$ 这 2 个参数，根据对辽宁省(锦州)、吉林省(扶余)和黑龙江省(海伦)的热量指数计算得到了东北 3 省热量指数预测模型中的 $\hat{a}$ 和 $\hat{u}$(表 1)。

表 1　东北各省热量指数逐月 GM(1，1)模型的参数

预报月	省份	模型中的参数		关联度	回代平均准确率 /%	预测平均准确率 /%
		$\hat{a}$	$\hat{u}$			
5	辽宁	0.001 7	433.876 8	0.969 8	98.2	96.8
	吉林	－0.005 2	420.461 2	0.929 7	95.6	97.7
	黑龙江	－0.016 1	359.858 0	0.918 4	94.7	93.2
6	辽宁	0.003 1	350.201 8	0.963 1	98.3	97.0
	吉林	－0.005 9	336.243 0	0.939 9	96.6	97.4
	黑龙江	－0.019 4	280.533 0	0.899 2	94.7	91.6
7	辽宁	0.008 3	268.542 4	0.939 2	98.0	95.7
	吉林	－0.002 2	259.855 0	0.941 9	97.5	98.9
	黑龙江	－0.019 3	203.360 1	0.882 4	95.3	94.0
8	辽宁	0.015 3	174.871 2	0.901 9	97.8	95.7
	吉林	0.001 0	173.448 8	0.904 3	97.5	97.4
	黑龙江	－0.025 9	119.140 8	0.813 6	95.1	96.8

由表 1 可见，模型的关联度都通过了显著性检验。从回代的准确率看，辽宁省的准确率最高，黑龙江省的准确率最小；从试报的准确率看，吉林省的准确率最高，黑龙江省的仍为最低。虽然模型的平均准确率低于逐步回归统计模型，但各个模型的平均准确率仍都在 91% 以上，完全可以应用该模型的预测结果指导农业生产。

4　结论与讨论

(1) GM(1，1)是开展时间序列环境要素变化趋势预测的有效方法之一。本文通过对东北地区热量指数的分析，建立了热量指数时间变化趋势的 GM(1，1)预测模型，各模型的平均预测精度虽低于逐步回归统计模型，但也都达 91% 以上，可以应用该模型的预测结果指导农业生产。

(2) GM 模型是以灰色模块为基础，以微分拟合法而建成的模型。因此，它由原点(现在时刻)向未来时刻呈喇叭状展开，即未来时刻越远，预测值灰区间就越大。这样，模型对系统的刻画将因时间的逐渐外推而逐渐失真，一般情况下，在模型使用 3 ~5 年之后，需要根据实况资料重新建模。

(3) 针对 GM 模型存在的问题，灰色系统理论提出了一系列调整和修正模型的方法。建立的GM(1，1)模型在没有达到所需要的预测精度的情况下，可以建立残差 GM(1，1)模型、残差序列周期修正 GM(1，1)模型、等维灰色(或新息)递补 GM(1，1)模型、LGM(1，1)长期预测模型等等，对原模型进行修正可以进一步提高预测精度，使得较短的原始数列也可以进行长期预测。

参考文献:

[1] 曹鸿兴，郑耀文，顾今. 灰色系统理论浅述[M]. 北京：气象出版社，1988：41 -48.

[2] 李习平. 基于 GM(1，1)理论的中国居民消费价格指数预测模型研究[J]. 全国商情(经济理论研究)，2009(4)：133 -134.

[3] 张睿，高焕文. 基于灰色 GM(1，1)的农业机械化水平预测模型[J]. 农业机械学报，2009，40(2)：91 -95.

[4] 罗振瑞，钟文静. 马尾松毛虫大发生灰色灾变趋势测报方法[J]. 安徽农学通报，2008，14(22)：83 -84.

[5] 李德庆. 利用 GM(1，1)模型群预测油田产油量[J]. 油气地质与采收率，2008，15(5)：82 -85.

[7] 胡丽敏，周新地，黄长军. 灰色模型 GM(1，1)在益阳市耕地预测中的应用[J]. 湖南城市学院学报：自然科学版，2008，17(3)：75 -78.

[8] 李文亮，张冬有，张丽娟. 黑龙江省低温冷害发生规律及预测研究[J]. 灾害学，2008，23(4)：30 -35.

[9] 程伟，刘国璧. 灰色系统理论在粮食产量预测中的应用[J]. 湖南工程学院学报：自然科学版，2008，18(3)：64 -67.

[10] 石娟，骆有庆，武海卫，等. 松材线虫(Bursaphelenchus xylophilus)入侵对马尾松(Pinus massoniana)林分生长的影响及相关生长模型[J]. 生态学报，2008，28(7)：3193 -3204.

[11] 张星，陈惠，周乐照. 福建省农业气象灾害灰色评价与预测[J]. 灾害学，2007，22(4)：43 -45，56.

[12] 胡小晖，延军平，欧维新. 1950 年以来广西洪涝灾害及趋势预测[J]. 灾害学，1999，14(4)：27 -31.

[13] 李彤，黄岁樑. 灰色 GM(1，2，M，N)在地面沉降预测中的应用[J]. 灾害学，2007，22(2)：56 -61.

[14] 袁嘉祖. 灰色系统理论及其应用[M]. 北京：科学出版社，1991.

[15] 傅立. 灰色系统理论及其应用[M]. 北京：科学技术文献出版社，1992.

[16] 祈宦，王颖，王昉. 灰色 - 马尔柯夫链预测棉花产量(单产)[J]. 安徽农业科学，2002，30(1)：152 -154.

Study on Forecasting Methods of Corn Heat Index in Northeastern China(Ⅲ)

——GM(1，1) Forecasting Model

Guo Jianping[1]，Chen Yueyi[1,2] and Zhuang Liwei[3]

(1. *Chinese Academy of Meteorological Sciences*，*Beijing* 100081，*China*；2. *Guangdong Meteorological Bureau*，*Guangzhou* 510080，*China*；3. *National Meteorological Center*，*Beijing* 100081，*China*)

Abstract：GM(1，1) is one of the effective methods for forecasting changing tendency of environmental factors. On the basis of corn heat index analysis in Northeastern China，the GM(1，1) forecasting model of corn heat index changing tendency is set up. The mean precision of each model is lower than that of stepwise regression models，but the mean precision is higher than 91%. The forecasting results of GM(1，1) can be used to instruct agricultural production.

Key words：northeastern China；corn；heat index；forecast；GM(1，1) Forecasting Model