

基于驱动因素控制的DFCGM(1, N) 及其拓展模型构建与应用

丁松^{1†}, 党耀国¹, 徐宁², 朱晓月¹

(1. 南京航空航天大学 经济与管理学院, 南京 211106; 2. 南京审计大学 管理科学与工程学院, 南京 211815)

摘要: 针对一类驱动因素具有复杂变化特征的系统行为预测问题, 将驱动因素序列对系统的作用函数引入经典GM(1, N)模型的灰色作用量, 构建驱动因素控制的DFCGM(1, N)模型及其拓展模型, 并探讨参数估计方法; 从白化信息充分和匮乏两个角度, 利用经验分析法和智能优化算法探索驱动因素控制参数的识别方法, 并给出建模预测步骤; 最后, 通过对我国粮食产量进行预测, 验证了模型的有效性和实用性, 表明所提出模型能够有效解决多驱动因素影响的系统预测问题。

关键词: 灰色预测; 驱动因素控制; DFCGM(1, N); 粮食产量预测

中图分类号: N941.5

文献标志码: A

Modeling and applications of DFCGM(1, N) and its extended model based on driving factors control

DING Song^{1†}, DANG Yao-guo¹, XU Ning², ZHU Xiao-yue¹

(1. College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; 2. College of Management Science and Engineering, Nanjing Audit University, Nanjing 211815, China)

Abstract: To solve the forecasting problems of system behavior affected by driving factors with a complicated changing patterns, a driving factor control grey model of N variables driving factor control grey model of N variables(DFCGM(1, N)) model based on driving factors control and its derived model are designed by introducing a driver function into the construction of the grey action item in the conventional GM(1, N) model, and subsequently the method of estimating the parameters is discussed. According to the sufficient and insufficient information about the driving factors, empirical analysis and intelligent optimization algorithms are used to explore the estimation of driving factors control parameters, respectively, and subsequently the modeling procedure is given with a flowchart. Finally, the effectiveness and practicality of the proposed model are verified by forecasting the grain production in China. Results show that the proposed model can effectively solve the forecasting problems on the system affected by multiple driving factors.

Keywords: grey forecasting; driving factors control; driving factor control grey model of N variables; forecasting grain production

0 引言

在经济社会系统中, 各组成要素间存在复杂多变、交织影响的变量关系, 某一系统特征序列的变化必然受到多种影响因素的作用, 而这种作用机制可能包含不确定、信息不完备的本质性特征。为了解决这一类系统行为序列的趋势预测问题, 邓聚龙教授在单

变量模型的基础上提出了多变量GM(1, N)模型^[1], 并且成功应用到经济社会的多个领域。在理论和方法层面, 后续学者对多变量预测模型的数乘变换等性质、参数优化、模型改进与拓展进行研究, 从而有效地提高了模型的建模精度, 拓宽了模型的适用范围。

肖新平等^[2]研究了数乘变换特征下GM(1, N)

收稿日期: 2017-03-07; 修回日期: 2017-08-15.

基金项目: 国家自然科学基金项目(71371098, 71771119, 71701101); 江苏省普通高校研究生科研创新计划项目(KYZZ16_0153); 南京航空航天大学博士学位论文创新与创优基金项目(BCXJ16-09); 江苏省社科基金重点研究项目(16GLA001); 江苏省高校自然科学研究项目(16KJD120001).

责任编委: 陈虹.

作者简介: 丁松(1992—), 男, 博士生, 从事灰色系统理论及其应用的研究; 党耀国(1964—), 男, 教授, 博士生导师, 从事灰色系统理论及其应用等研究.

†通讯作者. E-mail: dingsong1129@163.com

模型的参数估计、模型精度的前后变化情况,说明利用数乘变换技术可以减少矩阵漂移现象,优化模型建模效果. Guo等^[3]优化了GM(1, N)模型的背景值,并结合残差修正技术提出了全适性的GM(1, N)模型,提升了该模型对负数序列、正负交叉序列、正序列等多种数据类型的建模处理能力. Guo等^[4]结合自记忆性原理的优点,提出了自记忆性多变量预测模型. 针对传统模型模拟和预测精度不理想的问题, Zeng等^[5]引入线性修正项和灰作用项,克服传统模型的结构缺陷,提升了模型对变量间数据变化规律和相互依赖关系的描述能力,进而改善建模效果. 鉴于该模型在处理多因素影响下系统行为预测问题的有效性, Zeng等^[6]还用该优化模型预测北京机动车的数量,取得了较好的预测效果. 丁松等^[7]将驱动变量的趋势发展引入模型结构中,构建了基于发展趋势的TGM(1, N)模型. 为了提升对非线性多变量序列的处理能力,王正新^[8]提出了灰色GM(1, N)幂模型,并给出了幂指数的求解方法. 在此基础上, Ding等^[9]利用新的多变量模型对我国CO₂排放量进行了预测.

为了更准确描述系统行为与驱动因素间的真实关系,学者融入其他建模技术,不断拓宽多变量灰色预测模型的应用领域. Tien^[10]基于卷积积分技术,提出了GMC(1, N)模型,并提出了一系列改进模型^[11-12],以提升对多种特征序列的处理能力. Wang等^[13]将插值系数引入GMC(1, N)模型背景值中,利用非线性优化模型进行参数优化,并预测我国工业能源消费量. 张可等^[14]通过引入时滞控制项,并结合灰色扩维方法探索驱动项选择和时滞参数测算. 毛树华等^[15]根据多变量系统存在的时滞特性,利用分数阶累加生成算子,提出分数阶累加GM(1, N, τ)模型. 王正新^[16]通过引入时滞累积影响系数,构建多变量时滞GM(1, N)模型,并取得了良好的预测效果. Xie等^[17]将单变量离散模型拓展到多变量,并与传统GM(n, h)对比分析,研究其相关性质. 在此基础上, Ma等^[18]提出了递归离散多变量模型,并通过3个实例证明其适用性. Wu等^[19]提出了反向累加GM(1, N)模型,并融合滚动机理,对金砖国家的CO₂的排放量进行预测.

上述方法都有效地改善了GM(1, N)模型的建模效果,但通过考察现有模型可以发现,通过将驱动因素变量的一阶累加生成序列以线性组合的形式引入灰色作用量,默认了驱动变量自始至终对系统变量产生相同强度作用效果的假设. 而事实上,在实际应用中,各驱动因素在不同阶段对系统可能存在不同的作用效果. 系统运行过程中,主要驱动因素选择及其作用方式不尽相同,与主系统行为序列间存在错综复杂的关系,可能会出现 $X_i^{(1)}(i = 2, 3, \dots, N)$ 具

有较大变化幅度的情况,不对驱动因素作用机制进行分析必然会导致较大的模型误差. 因此,本文在传统GM(1, N)模型基础上,引入驱动因素作用机制控制项,提出DFCGM(1, N)模型及其拓展模型. 最后,结合我国粮食产量的实际数据,验证新模型的有效性,并对2016~2020年我国粮食产量进行预测建模.

1 驱动因素控制的GM(1, N)模型构建

1.1 模型定义及其参数估计

定义1^[1] 设原始序列为 $X_i^{(0)}$,其1-AGO序列为 $X_i^{(1)}$, $X_1^{(0)}$ 为系统行为序列, $X_i^{(0)}(i = 2, 3, \dots, N)$ 为驱动因素序列,称

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{i=2}^N b_i x_i^{(1)}(k) \quad (1)$$

为GM(1, N)模型. 其中: a 为系统发展系数, $b_i x_i^{(1)}(k)$ 为驱动项, b_i 为驱动系数, $Z_1^{(1)}$ 为 $X_1^{(1)}$ 的紧邻均值生成序列, $\hat{a} = [a, b_2, \dots, b_N]$ 为模型的参数列.

定义2^[1] 设 $\hat{a} = [a, b_2, \dots, b_N]$,称

$$\frac{dx_1^{(1)}(t)}{dt} + ax_1^{(1)}(t) = \sum_{i=2}^N b_i x_i^{(1)}(t) \quad (2)$$

为GM(1, N)模型的白化方程,或影子方程.

GM(1, N)模型能够反映驱动因素对系统行为变量的同步作用效果,在已知驱动变量变化趋势信息及变动幅度较小的条件下,能够对系统行为变量进行预测. 但是,该模型未考虑驱动因素对主系统的作用机制,默认各驱动因素在系统运行过程中始终产生相同的作用强度. 而现实生活中,随着外界作用介质变化以及系统自身的演化发展,系统运行过程中各驱动因素的作用时效和强度不尽相同,发挥主导作用的驱动因素也可能产生变化,会出现 $X_i^{(1)}(i = 2, 3, \dots, N)$ 产生大幅变化的情况,从而不适宜构建传统模型. 因此,本文在GM(1, N)模型基础上,引入驱动因素作用机制控制项,提出基于驱动因素控制的GM(1, N)模型,并引出其派生模型.

定义3 设原始序列为 $X_i^{(0)}$,其1-AGO序列为 $X_i^{(1)}$, $X_1^{(0)}$ 为系统行为序列, $X_i^{(0)}(i = 2, 3, \dots, N)$ 为驱动因素序列,称

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{i=2}^N b_i d_i(k) x_i^{(1)}(k) \quad (3)$$

为基于驱动因素控制的GM(1, N)模型,记为DFCGM(1, N). 其中: $d_i(k) = \mu(k - \gamma_{i,1}) - \mu(k - \gamma_{i,2}) (0 \leq \gamma_{i,1} \leq \gamma_{i,2})$ 为驱动因素 $X_i(i = 2, 3, \dots, N)$ 的控制项, $\mu(k)$ 为阶跃函数, $\gamma_{i,1}$ 和 $\gamma_{i,2}$ 分别为驱动因素 $X_i(i = 2, 3, \dots, N)$ 的开始和结束作用时间.

在DFCGM(1, N)模型中,背景值为紧邻均值生成序列, $\sum_{i=2}^N b_i d_i(k) x_i^{(1)}(k)$ 为驱动因素项. 若不考虑驱

动因素作用机制,则该模型退化为GM(1, N)模型.

定理1 当 $\gamma_{i,1} = 0$ 且 $\gamma_{i,2} \rightarrow +\infty$ 时,GM(1, N)与DFCGM(1, N)模型等价.

证明 当 $\gamma_{i,1} = 0$ 且 $\gamma_{i,2} \rightarrow +\infty$ 时, $d_i(k) = \mu(k - \gamma_{i,1}) - \mu(k - \gamma_{i,2}) = \mu(k)$,即对于任意的驱动因素 $X_i^{(0)}(i = 2, 3, \dots, N)$, $d_i(k)$ 为单位阶跃函数.又因为 $\mu(k) = \begin{cases} 1, & k \geq 0 \\ 0, & k < 0 \end{cases}$,则式(3)可以简化为

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{i=2}^N b_i d_i(k) x_i^{(1)}(k) = \sum_{i=2}^N b_i x_i^{(1)}(k). \quad \square$$

定理1表明:系统运行过程中,若所有驱动因素从开始至结束始终对系统行为序列产生强度相当的影响,则两种模型是等价的;而随着系统的演化发展,对系统产生主导作用的驱动因素序列在系统运行过程中发生变化,此时应选择能反映驱动因素变化规律的DFCGM(1, N)模型.因此,本文模型是对传统模型的传承,是在驱动因素作用机制信息不断白化条件下对传统多变量模型的进一步拓展与丰富.

定理2 若 $\hat{b} = (a, b_2, \dots, b_N)^T$ 为参数列,且

$$Y = [x_1^{(0)}(2) \ x_1^{(0)}(3) \ \dots \ x_1^{(0)}(n)]^T,$$

$B =$

$$\begin{bmatrix} -z_1^{(1)}(1) & d_2(2)x_2^{(1)}(2) & \dots & d_N(2)x_N^{(1)}(2) \\ -z_1^{(1)}(2) & d_2(3)x_2^{(1)}(3) & \dots & d_N(3)x_N^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ -z_1^{(1)}(n-1) & d_2(n)x_2^{(1)}(n) & \dots & d_N(n)x_N^{(1)}(n) \end{bmatrix},$$

则参数列的最小二乘法估计满足:

- 1) 当 $n = N + 1$ 时, $\hat{b} = B^{-1}Y, |B| \neq 0$;
- 2) 当 $n > N + 1$ 时, $\hat{b} = (B^T B)^{-1} B^T Y, |B^T B| \neq 0$;
- 3) 当 $n < N + 1$ 时, $\hat{b} = B^T (B^T B)^{-1} Y, |B^T B| \neq 0$.

证明过程请参考文献[8]和文献[16].

定义4 设 $\hat{b} = (a, b_2, \dots, b_N)^T$,称

$$\frac{dx_1^{(1)}(t)}{dt} + ax_1^{(1)}(t) = \sum_{i=2}^N b_i d_i(t) x_i^{(1)}(t)$$

为DFCGM(1, N)模型的白化方程.

定理3 设 $X_i^{(0)}, X_i^{(1)}, B, Y$ 如以上定义所示,有:

- 1) DFCGM(1, N)模型白化微分方程的解为

$$x_1^{(1)}(t) = e^{-at} \left\{ \sum_{i=2}^N \int b_i d_i(t) x_i^{(1)}(t) e^{at} dt + c \right\}, \quad (4)$$

其中 c 为待定常数.将初始条件 $x_i^{(1)}(0)$ 代入式(4),得

$$c = x_1^{(1)}(0) - \sum_{i=2}^N \int b_i d_i(0) x_i^{(1)}(0) e^{at} dt, \quad (5)$$

所以 $x_1^{(1)}(t)$ 可以进一步表达为

$$x_1^{(1)}(t) = e^{-at} \sum_{i=2}^N \int b_i d_i(t) x_i^{(1)}(t) e^{at} dt + e^{-at} \left\{ x_1^{(1)}(0) - t \sum_{i=2}^N b_i d_i(0) x_i^{(1)}(0) \right\}. \quad (6)$$

2) 当驱动因素序列 $X_i^{(1)}(i = 2, 3, \dots, N)$ 变化幅度较小时,模型的驱动项 $\sum_{i=2}^N b_i d_i(k) x_i^{(1)}(k)$ 可以视为灰常量,DFCGM(1, N)模型灰色微分方程的近似时间响应函数序列为

$$\hat{x}_1^{(1)}(k) = e^{-a(k-1)} \left\{ x_1^{(1)}(1) - \frac{1}{a} \sum_{i=2}^N b_i d_i(k) x_i^{(1)}(k) \right\} + \frac{1}{a} \sum_{i=2}^N b_i d_i(k) x_i^{(1)}(k). \quad (7)$$

其1-IAGO为 $\hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(1)}(k)$.定理2和定理3分别给出了DFCGM(1, N)模型的参数估计和模拟值计算公式,当主系统行为和驱动因素作用机制已知时,可通过上述定理进行计算.

1.2 DFCGM(1, N)模型的拓展模型

本文模型与GM(1, N)模型能进行预测的前提是驱动因素项为灰常量,而该条件在实际应用背景下很难达到.驱动因素序列的波动性及与主行为序列间作用机制的复杂性可能会导致驱动因素变化幅度较大,造成模型预测效果不理想.因此,探讨DFCGM(1, N)的拓展形式,减少运算工作量,是比较有意义的工作.

定理4 从DFCGM(1, N)的灰色微分方程可以推导出其派生模型,即DFCGM(1, N, $x^{(1)}$)模型

$$\hat{x}_1^{(0)}(k) = \sum_{i=2}^N \frac{b_i}{1 + 0.5a} d_i(k) x_i^{(1)}(k) - \frac{a}{1 + 0.5a} x_1^{(1)}(k-1). \quad (8)$$

证明 灰导数 $x_1^{(1)}(k)$ 的背景值为

$$z_1^{(1)}(k) = 0.5x_1^{(1)}(k) + 0.5x_1^{(1)}(k-1) = x_1^{(1)}(k-1) + 0.5x_1^{(0)}(k),$$

将其代入灰色微分方程(3)可得

$$x_1^{(0)}(k) + a[x_1^{(1)}(k-1) + 0.5x_1^{(0)}(k)] = \sum_{i=2}^N b_i d_i(k) x_i^{(1)}(k), \quad (9)$$

进一步化简可得

$$x_1^{(0)}(k) = \sum_{i=2}^N \frac{b_i}{1 + 0.5a} d_i(k) x_i^{(1)}(k) - \frac{a}{1 + 0.5a} x_1^{(1)}(k-1). \quad (10)$$

定理得证. \square

定理4仅是本文模型的一个拓展形式,能实现在驱动因素变动幅度较大的情形下模型的建模预测,在该拓展模型的基础上还能进行更多形式的推导.

2 驱动因素控制项识别及建模步骤

在实际建模过程中,各驱动因素的始末作用时期和强度是未知的,需要通过不断增加白化信息对各参数进行估计.

2.1 驱动因素控制项参数识别方法

1) 白化信息充分. 若已知系统行为状态或驱动因素作用时段及状态,则可以根据经验给出驱动因素作用信息. 依据各类型白化信息,将主系统序列和驱动序列分为若干子序列,通过分析系统序列子序列与各驱动序列子序列的相关性,进一步确定主导驱动因素的作用机制.

定义5 设 $P = \{P_1, P_2, \dots, P_m | 1 \leq P_i \leq n\}$, P_i 为整数,将主系统行为序列和驱动因素序列 $X_i (i = 1, 2, \dots, N)$ 划分为 $m + 1$ 个子序列,分别记为

$$\begin{aligned} X_i^1 &= \{X_i^1, X_i^2, \dots, X_i^{P_1}\}, \\ X_i^j &= \{X_i^{P_{j-1}+1}, X_i^{P_{j-1}+2}, \dots, X_i^{P_j}\}, \\ X_i^{m+1} &= \{X_i^{P_m+1}, X_i^{P_m+2}, \dots, X_i^n\}. \end{aligned}$$

利用灰色关联分析方法计算主系统行为序列 X_1 与驱动因素序列 $X_i (i = 2, 3, \dots, N)$ 的第 j 个子序列的关联度,记为 ε_i^j . 设关联度阈值为 ε ,若 $\varepsilon_i^j \geq \varepsilon$,则称 X_i^j 为第 j 段的主驱动因素序列. 进而可得在第 j 段的驱动因素控制函数为

$$d_i(k) = \mu(k - p_{j-1} - 1) - \mu(k - p_j).$$

若 X_i 为第 $j, j + 1, \dots, j + q$ 时段为主系统行为序列的主驱动因素序列,则

$$d_i(k) = \mu(k - p_{j-1} - 1) - \mu(k - p_{j+q}).$$

定义5根据掌握充分的白化信息,利用灰色关联分析方法,确定各时段中对主系统行为序列影响作用较大的主驱动因素,进而利用定义3获得驱动因素控制函数,然后用原模型或拓展模型进行建模预测. 当主系统行为信息较为充分,驱动因素作用机制明确时,该方法行之有效,具有实际应用价值. 然而,针对相对匮乏的白化信息,该方法将难以识别驱动因素控制项,需要进一步采取智能优化算法进行参数识别.

2) 白化信息匮乏. 设各驱动因素控制项的待优化参数序列为 $\gamma = \{(\gamma_{2,1}, \gamma_{2,2}), (\gamma_{3,1}, \gamma_{3,2}), \dots, (\gamma_{N,1}, \gamma_{N,2})\}$, $\hat{x}_1^{(0)}$ 为模拟值,以建模的平均相对误差最小化为目标,以模型参数之间的关系为约束条件,构建以下非线性优化模型,借助粒子群等智能算法进行参数优化选择.

i) 驱动因素项 $X_i^{(1)} (i = 2, 3, \dots, N)$ 变化幅度较小时,采用式(7)作为时间响应函数,通过以下非线性

规划模型求解驱动控制项时间参数:

$$\begin{aligned} \min \text{avg}(Q(\gamma)) &= \sum_{k=2}^n \frac{|\hat{x}_1^{(0)}(k) - x_1^{(0)}(k)|}{x_1^{(0)}(k)}, \\ \text{s.t.} \quad &\begin{cases} \hat{x}_1^{(0)}(k) = \hat{x}_1^{(1)}(k) - \hat{x}_1^{(1)}(k-1), \\ \hat{b} = (a, b_2, \dots, b_N)^T, \\ 0 \leq \gamma_{i,1} \leq \gamma_{i,2}, \\ i = 2, 3, \dots, N, k = 2, 3, \dots, n. \end{cases} \end{aligned} \quad (11)$$

ii) 驱动因素项 $X_i^{(1)} (i = 2, 3, \dots, N)$ 变化幅度较大时,采用式(8)作为时间响应函数,通过以下非线性规划模型求解驱动控制项时间参数:

$$\begin{aligned} \min \text{avg}(Q(\gamma)) &= \sum_{k=2}^n \frac{|\hat{x}_1^{(0)}(k) - x_1^{(0)}(k)|}{x_1^{(0)}(k)}, \\ \text{s.t.} \quad &\begin{cases} \hat{x}_1^{(0)}(k) = \\ \sum_{i=2}^N \frac{b_i}{1 + 0.5a} d_i(k) x_i^{(1)}(k) - \frac{ax_1^{(1)}(k-1)}{1 + 0.5a}, \\ \hat{b} = (a, b_2, \dots, b_N)^T, \\ 0 \leq \gamma_{i,1} \leq \gamma_{i,2}, \\ i = 2, 3, \dots, N, k = 2, 3, \dots, n. \end{cases} \end{aligned} \quad (12)$$

通过以上两个优化问题,借助智能优化算法可以求得驱动因素控制项的时间参数,一旦求解出 γ 最优解,预测模型的结构参数 $\hat{b} = (a, b_2, \dots, b_N)^T$ 即可确定,进而可利用式(7)或(8)进行模拟和预测.

2.2 DFCGM(1, N) 建模步骤

Step 1: 根据相关资料研究和专家调研,分析影响主系统行为序列的影响因素,初步确定驱动因素 $X_i^{(0)}, i = 2, 3, \dots, N$,根据定义3构建DFCGM(1, N)模型.

Step 2: 若系统运行的白化信息较为充分,则根据定义5识别驱动因素控制项的时间参数和确定主要驱动因素;若系统行为信息未知,则根据式(11)或(12)进行参数优化,通过智能算法获得驱动因素控制项参数.

Step 3: 根据定理2求解模型参数,再依据定理3或者拓展模型(8)计算主系统行为序列的模拟值,并进行精度效果检验. 若误差超出许可范围,则转入Step 1,重新分析驱动因素及控制参数.

Step 4: 对系统行为进行预测时,在各主驱动因素数值未知的情况下,需先采用GM(1, 1)或其他时间序列预测模型分别预测各主驱动因素的预测值;若已知各主驱动因素数值,则直接代入模型.

Step 5: 根据驱动因素预测值,采用DFCGM(1, N)模型或其拓展模型预测系统行为序列 $X_1^{(0)}$,最后分析预测结果的合理性及意义.

3 实例分析

粮食生产在我国国民经济发展中占有举足轻重的地位,尤其是新常态背景下我国经济下行压力加大而稳定粮食生产,提升粮食产量预测的准确性,有利于实现稳定经济增长的基本面.但由于粮食生产受到复杂多变的气候、地理环境及农业技术等多种因素变量的影响,我国粮食产量呈现复杂多变、波动发展的形态特征,因此,如何准确预测我国未来粮食产量成为学界比较关心的话题.为此,本文在充分挖掘我国粮食产量影响因素作用机制的基础上,建立DFCGM(1, N)模型进行模拟和预测.

为了检验本文模型的效果,以2003~2012年数

据作为建模数据,2013~2015年数据作为模型预测效果检验数据,数据来源为《中国统计年鉴》.通过对历史文献资料的研究发现,对粮食产量产生影响的因素主要为耕地灌溉面积、农业机械总动力、化肥施用量及粮食播种面积.各因素的测量值如表1所示,分别记粮食产量为 X_1 、各影响因素为 $X_2 \sim X_5$.为了避免不同指标数据数量级差异带来的模型建模误差,在建模前需对数据进行初值化处理,所有建模均采用变换后的数据,获得模拟预测结果后,再进行数据还原处理.根据定义3建立DFCGM(1, N)模型为

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{i=2}^5 b_i d_i(k) x_i^{(1)}(k). \quad (13)$$

表1 2003~2015年我国粮食产量及影响因素

年份	粮食产量 X_1 / 万吨	灌溉面积 X_2 / 千公顷	机械总动力 X_3 / 万千瓦	化肥施用量 X_4 / 万吨	播种面积 X_5 / 千公顷
2003	43 069.50	540 14.23	60 386.54	4 411.56	99 410.37
2004	46 946.90	54 478.42	64 027.91	4 636.58	101 606.03
2005	48 402.20	55 029.34	68 397.85	4 766.22	104 278.38
2006	49 804.20	55 750.50	72 522.12	4 927.69	104 958.00
2007	50 160.28	56 518.34	76 589.56	5 107.83	105 638.36
2008	52 870.92	58 471.68	82 190.41	5 239.02	106 792.65
2009	53 082.08	59 261.40	87 496.10	5 404.40	108 985.75
2010	54 647.71	60 347.70	92 780.48	5 561.68	109 876.09
2011	57 120.80	61 681.56	97 734.66	5 704.24	110 573.02
2012	58 957.97	62 490.52	102 558.96	5 838.85	111 204.59
2013	60 193.84	63 473.30	103 906.75	5 911.86	111 955.56
2014	60 702.60	64 540.00	108 056.60	5 995.90	112 723.00
2015	62 143.90	65 873.00	111 728.10	6 022.60	113 343.00

通过上述分析可知,粮食产量序列和各驱动因素序列的白化信息比较充分,可以利用定义5识别主驱动因素及其作用控制项函数.因此,以2007年为分界年,分别研究2003~2007年、2008~2012年各驱动因素对粮食产量的影响程度,计算各驱动因素与粮食产量的关联度如表2所示.为了提高两个阶段驱动因素的区分度,本文选择常用的关联度阈值 $\varepsilon = 0.75$.由定义5可知,2003~2007年间对粮食产量影响较大的驱动因素为 X_3 和 X_4 ,而2008~2012年间的主驱动因素为 X_4 .驱动因素作用机制分析结果表明,2003~2007年间农业机械化和化肥使用程度的提升是粮食产量增加的主要驱动因素,而2008年之后,农业机械化达到一定规模后,随着农业机械化程度达到一定水平后会出现边际效应递减现象;而化肥的使用一直是我国粮食产量保持增长趋势的基础,随着化肥技术和品质的不断提升,土地贫瘠现象的加重,粮食产量对其依赖程度也不断提高,这与我国当前农业种植现状比较吻合.根据以上分析,可以得到各驱动因素的控制函数为

$$\begin{aligned} d_3(k) &= \mu(k-1) - \mu(k-6), \\ d_4(k) &= \mu(k-1), \\ d_2(k) &= d_5(k) = 0. \end{aligned}$$

表2 各驱动因素与粮食产量的关联度计算结果

年份	灌溉面积 X_2 / 千公顷	机械总动力 X_3 / 万千瓦	化肥施用量 X_4 / 万吨	播种面积 X_5 / 千公顷
2003~2007	0.496 4	0.804 1	0.832 7	0.538 3
2008~2012	0.623 5	0.532 4	0.857 5	0.572 9

式(13)所示模型转化为

$$\begin{aligned} x_1^{(0)}(k) + az_1^{(1)}(k) &= \\ &[\mu(k-1) - \mu(k-6)]b_3x_3^{(1)}(k) + \\ &\mu(k-1)b_4x_4^{(1)}(k). \end{aligned} \quad (14)$$

由原始数据可以发现, X_3 和 X_4 数据变化幅度较大时驱动项不能视为灰常量,为了避免由此导致的误差,选择建立DFCGM(1, N)拓展模型.根据定理2和定理4求解模型参数,可得

$$a = 1.9709, b_3 = 0.0185, b_4 = 2.0052.$$

即本文模型预测模型为

$$\begin{aligned} \hat{x}_1^{(0)}(k) &= 0.0093d_3(k)x_3^{(1)}(k) + 1.0099d_4(k)x_4^{(1)}(k) - \\ &0.9927x_1^{(1)}(k-1). \end{aligned} \quad (15)$$

根据上述拓展模型模拟2003~2012年粮食产量,结果见表3.将2013~2015年的 X_3 和 X_4 数据代入式(15)进行粮食产量预测.若 X_3 和 X_4 数据未知,

表3 我国粮食产量模拟预测结果

年份	原始数据	DFCGM(1, N) 模型		GM(1, N) 模型		DGM(1, N) 模型		ARMA 模型	
		模拟值	误差 / %	模拟值	误差 / %	模拟值	误差 / %	模拟值	误差 / %
2003	43 069.50	43 069.50	0.00	43 069.50	0.00	43 069.50	0.00	42 793.85	0.64
2004	46 946.90	46 459.07	1.03	43 465.74	7.42	46 928.53	0.04	34 042.50	27.49
2005	48 402.20	46 838.08	3.25	57 583.92	18.97	60 430.82	24.83	38 140.77	21.20
2006	49 804.20	47 367.84	4.86	54 659.50	9.75	53 974.70	8.41	38 240.94	23.22
2007	50 160.28	48 315.37	3.71	52 010.73	3.69	55 727.63	11.07	40 710.03	18.84
2008	52 870.92	50 184.58	5.12	55 128.96	4.27	58 484.07	10.58	43 278.45	18.14
2009	53 082.08	50 968.45	3.95	55 970.39	5.44	59 155.96	11.48	45 775.54	13.76
2010	54 647.71	53 147.76	2.76	57 115.10	4.52	60 624.63	10.92	48 119.11	11.95
2011	57 120.80	55 133.27	3.46	58 760.51	2.87	62 554.14	9.53	50 453.73	11.67
2012	58 957.97	56 033.42	4.97	58 957.97	3.43	63 673.95	7.99	52 724.54	10.57
模拟值平均相对误差 / %		3.31		6.71		9.48		15.75	

年份	原始数据	DFCGM(1, N) 模型		GM(1, N) 模型		DGM(1, N) 模型		ARMA 模型	
		预测值	误差 / %	预测值	误差 / %	预测值	误差 / %	预测值	误差 / %
2013	60 193.84	58 273.03	3.19	62 577.25	3.96	64 940.19	7.85	54 890.28	8.81
2014	60 702.60	59 522.05	1.94	64 103.94	5.60	66 464.85	9.53	57 124.45	5.89
2015	62 143.90	60 986.41	1.86	66 878.71	7.62	68 467.58	10.40	59 425.13	4.37
预测值平均相对误差 / %		2.33		5.73		9.26		6.36	

则可利用GM(1, 1)模型或其他时间序列模型进行驱动因素预测, 然后对系统行为进行预测; 这里是为了验证模型的有效性, 2013~2015年 X_3 和 X_4 的数据已知, 直接代入式(15)进行建模预测即可. 为了对比模型的性能, 本文还选取了GM(1, N)、DGM(1, N)和ARMA模型作为对比模型, 所得模型计算结果如表3所示. 从各模型的建模效果来看, 本文模型在模拟和预测阶段均取得了较好的精度, 在充分分析驱动因素对粮食产量作用机制的基础上, 进一步提高了建模精度, 两项误差分别为3.31%和2.33%, 实现了高精度的平稳模拟和预测. 鉴于本文在通过初值化处理建模数据后, 数据变动幅度不大, 使得传统模型也取得了不错的预测效果, 模拟误差和预测误差分别为6.71%和5.73%. 但对于影响因素变幅较大的数据序列, GM(1, N)模型难以适应, 不满足其建模假设条件, 这些在以往文献中已得到验证^[13], 而且GM(1, N)模型不能很好地反映不同阶段主驱动因素的变化情况, 因此建模效果劣于本文模型. 对于DGM(1, N)模型, 虽然能够在一定程度上避免模型间的跳跃性误差, 但其未充分利用驱动信息, 忽视了各阶段主驱动因素的变化情况, 因此建模误差较大, 模拟和预测误差分别为9.48%和9.26%, 不适合用于粮食产量预测. 而作为经典统计模型, ARMA模型模拟阶段平均误差高达15.75%, 同时其平均预测误差也相对较高, 为6.36%, 劣于本文优化模型.

另外, 由图1的误差分布图可以看出, GM(1, N)模型精度波动大于本文模型, 离散多变量模型次之, 预测效果不稳定. 而ARMA模型, 误差波动程度最大, 远高于本文模型, 因此模拟预测效果最差. 综上所述, 对比3种模型模拟预测效果和误差分布图, 通

过分析粮食产量驱动因素的作用机制, 能够持续降低建模和预测误差, 本文方法更适合用于我国粮食产量的未来预测. 利用GM(1, 1)模型对主驱动因素的2016~2020年数值进行预测, 并将其代入本文提出的驱动控制的多变量预测模型, 其对2016~2020年我国粮食产量的预测结果见表4. 结果表明, 未来5年, 我国粮食产量将继续保持稳定增长, 预计到2020年突破6.5亿t, 继续为我国经济社会发展提供基本保证. 考虑到未来人口增长对粮食需求的不断增加, 粮食播种面积相对减少, 以及由化肥过度使用而带来的环境问题日益突出, 未来我国粮食增长将面临较大的挑战.

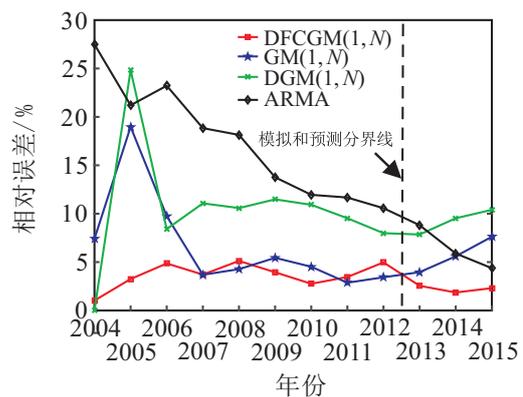


图1 4种模型的平均相对误差曲线图

表4 2016~2020年我国粮食产量预测

年份	2016	2017	2018	2019	2020
粮食产量/万t	62 709.19	63 226.03	63 785.93	64 431.97	65 207.22

4 结论

针对多变量系统中影响因素作用强度及时效差异性, 本文基于多变量灰色系统建模方法提出

了基于驱动因素控制的GM(1, N)模型,并给出了参数估计方法,为解决含多因素影响的系统行为预测提供了一种有效工具.在驱动因素序列变化幅度较小时,可采用白化方程近似解进行预测建模,反之则可采用拓展模型进行建模.当系统白化信息充分时,分析驱动因素的内部作用机制,识别主驱动因素并进行建模;反之,则借助智能算法,对驱动项时间控制参数方程进行最小平均误差条件下目标优化,进而对系统行为特征序列进行预测.实例研究表明,本文模型能够有效描述影响我国粮食产量的各驱动因素的作用机制,并能显著改善对我国粮食产量的模拟和预测效果.对于未来工作,还可以从两方面展开深入研究:1)鉴于文章篇幅所限,本文只列举了驱动因素白化信息充分条件下DFCGM(1, N)模型建模案例,对于白化信息匮乏条件下的DFCGM(1, N)模型案例研究以后可以展开广泛讨论,这样的案例在实际生活中比较常见,具有较高的研究价值;2)本文仅对驱动因素作用时间段及主驱动因素进行研究,未考虑不同时段内主驱动因素的作用强度变化,未来可以对这方面进行更深入研究.

参考文献(References)

- [1] Deng J L. The control problem of grey systems[J]. System Control Letter, 1982, 1(5): 288-294.
- [2] 肖新平, 毛树华. 灰预测与决策方法[M]. 北京: 科学出版社, 2013: 272-277.
(Xiao X P, Mao S H. Grey forecasting and decision making method[M]. Beijing: Science Press, 2013: 272-277.)
- [3] Guo H, Xiao X P, Forrest J. A research on a comprehensive adaptive grey prediction model CAGM(1, N)[J]. Applied Mathematics and Computation, 2013, 225(1): 216-227.
- [4] Guo X J, Liu S F, Wu L F, et al. A multi-variable grey model with a self-memory component and its application on engineering prediction[J]. Engineering Applications of Artificial Intelligence, 2015, 42: 82-93.
- [5] Zeng B, Luo C M, Liu S F, et al. Development of an optimization method for the GM(1, N) model[J]. Engineering Applications of Artificial Intelligence, 2016, 55: 353-362.
- [6] Zeng B, Luo C M, Li C, et al. A novel multi-variable grey forecasting model and its application in forecasting the amount of motor vehicles in Beijing[J]. Computers & Industrial Engineering, 2016, 101: 479-489.
- [7] 丁松, 党耀国, 徐宁. 基于驱动变量增长趋势的TGM(1, N)预测模型[J]. 系统工程, 2015, 33(3): 149-153.
(Ding S, Dang Y G, Xu N. TGM(1, N) prediction model based on trends of driving variables[J]. Systems Engineering, 2015, 33(3): 149-153.)
- [8] 王正新. 灰色多变量GM(1, N)幂模型及其应用[J]. 系统工程理论与实践, 2014, 34(9): 2357-2363.
(Wang Z X. Grey multivariable power model GM(1, N) and its application[J]. Systems Engineering—Theory & Practice, 2014, 34(9): 2357-2363.)
- [9] Ding S, Dang Y G, Li X M, et al. Forecasting Chinese CO₂ emissions from fuel combustion using a novel grey multivariable model[J]. J of Cleaner Production, 2017, 162: 1527-1538.
- [10] Tien T L. The indirect measurement of tensile strength of material by the grey prediction model GMC(1, N)[J]. Measurement Science Technology, 2005, 16(6): 1322-1328.
- [11] Tien T L. Forecasting CO₂ output from gas furnace by grey model IGMC(1, N)[J]. J of the Chinese Society of Mechanical Engineers, 2010, 31(1): 55-65.
- [12] Tien T L. The indirect measurement of tensile strength by the new model FGMC(1, N)[J]. Measurement, 2011, 44(10): 1884-1897.
- [13] Wang Z X, Hao P. An improved grey multivariable model for predicting industrial energy consumption in China[J]. Applied Mathematical Modelling, 2016, 40(11): 5745-5758.
- [14] 张可, 曲品品. 时滞多变量离散灰色模型及其应用[J]. 系统工程理论与实践, 2015, 35(8): 2092-2103.
(Zhang K, Qu P P. Delay multi-variables discrete grey model and its application[J]. Systems Engineering—Theory & Practice, 2015, 35(8): 2092-2103.)
- [15] 毛树华, 高明运, 肖新平. 分数阶累加时滞GM(1, N, τ)模型及其应用[J]. 系统工程理论与实践, 2015, 35(2): 430-436.
(Mao S H, Gao M Y, Xiao X P. Fractional order accumulation time-lag GM(1, N, τ) model and its application[J]. Systems Engineering—Theory & Practice, 2015, 35(2): 430-436.)
- [16] 王正新. 多变量时滞GM(1, N)模型及其应用[J]. 控制与决策, 2015, 30(12): 2298-2304.
(Wang Z X. Multivariable time-delayed GM(1, N) model and its application[J]. Control and Decision, 2015, 30(12): 2298-2304.)
- [17] Xie N M, Liu S F. Discrete grey forecasting model and its optimization[J]. Applied Mathematical Modelling, 2009, 33(2): 1173-1186.
- [18] Ma X, Liu Z. Research on the novel recursive discrete multivariate grey prediction model and its applications[J]. Applied Mathematical Modelling, 2016, 40(8): 4876-4890.
- [19] Wu L, Liu S, Liu D, et al. Modelling and forecasting CO₂ emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multi-variable grey model[J]. Energy, 2015, 79: 489-495.

(责任编辑: 孙艺红)