

基于 Logistic 回归、Markov 过程和改进灰自助法的导弹备件需求预测

赵建忠¹, 徐廷学², 李海军², 尹延涛³

1. 海军航空工程学院研究生管理大队, 山东烟台 264001
2. 海军航空工程学院兵器科学与技术系, 山东烟台 264001
3. 海军航空工程学院科研部, 山东烟台 264001

摘要 为了提高预测间断性需求导弹备件的精度, 提出一种基于 Logistic 回归、Markov 过程和改进灰自助法的组合预测模型。将样本序列划分为解释变量序列和自相关序列, 对解释变量采用 Logistic 回归模型预测提前期非零需求发生概率, 对自相关序列采用 Markov 过程估计提前期非零需求发生概率, 把这两方面组合得到提前期非零需求发生概率, 再运用改进灰自助法进行需求分布确定, 得到最终的提前期需求。改进灰自助法先进行 Bootstrap 抽样, 进行 GM(1,1) 二次数据拟合, 既克服了 Bootstrap 法在小子样下的重复抽样问题, 又克服了 Bootstrap 法在小子样下仿真结果不可信的问题。实例表明, 提出的组合预测方法降低了预测误差, 说明了该方法的有效性、可行性和实用性。

关键词 组合预测; 间断性需求; 自相关序列; Logistic 回归; Markov 过程; 改进灰自助法

中图分类号 TP391.9

文献标志码 A

doi 10.3981/j.issn.1000-7857.2013.16.011

Demand Forecasting of Missile Spare Parts Based on Logistic Regression, Markov Process and Improved Grey Bootstrap Method

ZHAO Jianzhong¹, XU Tingxue², LI Haijun², YIN Yantao³

1. Graduate Students' Brigade, Naval Aeronautical and Astronautical University, Yantai 264001, Shandong Province, China
2. Department of Ordnance Science and Technology, Naval Aeronautical and Astronautical University, Yantai 264001, Shandong Province, China
3. Department of Scientific Research, Naval Aeronautical and Astronautical University, Yantai 264001, Shandong Province, China

Abstract In order to enhance the forecasting accuracy of intermittent demands of missile spare parts, a combined forecasting model based on the logistic regression, Markov process and the improved improved grey Bootstrap method is proposed. This model splits the sample series into the explanatory series and the auto-correlated series. The probabilities of nonzero demands for the explanatory series in the lead time is estimated by a Logistic regression model, the probabilities of nonzero demands for the auto-correlated series in the lead time is estimated by the Markov process, and they are combined to obtain the probabilities of nonzero demands in the lead time. Finally the demand distribution is determined by the improved grey Bootstrap method, where, the bootstrap sampling is made, and the data are matched by GM(1,1). Based on the principle of the grey bootstrap, the resample method is improved to avoid the bootstrap being repeatedly resampled in a small sample case, and the GM(1,1) twice data fitting model is used to solve the problem of the credibility of the bootstrap's simulated result in a small sample case. Experimental results show that the combined forecasting model can significantly reduce the prediction errors and the method is effective, feasible and practical for forecasting the demands of missile spare parts.

Keywords combined forecasting; intermittent demand; auto-correlated series; Logistic regression; Markov process; improved grey Bootstrap method

收稿日期: 2012-12-07; 修回日期: 2013-03-20

作者简介: 赵建忠, 工程师, 研究方向为武器装备综合保障理论与技术, 电子信箱: zjznavy@163.com

0 引言

导弹装备故障机理十分复杂,其备件需求大都呈现“短周期、小样本”的特点,而且相当一部分关重件的备件属于间断性需求^[1],传统的连续性需求预测方法不太适合间断性需求预测^[2]。目前,研究间断性需求的预测方法主要有指数平滑法、Croston法、自助(Bootstrap)法以及Markov Bootstrap(MB)法等^[3-5],但成果还不多。它们能够预测需求服从某种分布或需求序列存在自相关性等一些特殊的间断需求,都没有考虑影响需求的因素,具有一定的局限性,且当需求样本比较少时准确性不高。

统计表明,由备件可靠性产生的需求通常是平稳需求,需求序列具有自相关性;维修方式、使用环境等因素产生的需求是非平稳的,引起非平稳需求的因素在统计学中也称为解释变量^[6]。间断性需求是这两类因素的综合体现。因此,本文提出一种基于组合预测的模型,从解释变量序列和自相关序列两个方面进行组合预测,然后整合预测结果,从而达到提高预测准确性的目的。

1 组合预测的基本思想

导弹备件关重件需求的数学描述为:设 t 阶段实际发生的需求为 $x(t)(t=1,2,\dots,n)$,其中有多阶段的需求为 0,当需求为非零时,其数值的大小又具有很大的随机性。预测的目的就是估计提前期 L 内的总需求,称为提前期需求量,表示为

$$Y_{LMB} = \sum_{t=n+1}^{n+L} x_t$$

首先将需求时间序列 $\{x_1, x_2, \dots, x_n\}$ 转换成需求是否发生的时间序列 $\{y_i\}(i=1,2,\dots,n)$, $y_i=0$ 表示没有需求; $y_i=1$ 表示发生非零需求。然后区分非零需求发生的原因,需求由自相关原因引起的划入“自相关序列”,需求由解释变量原因引起的划入“解释变量序列”。

借鉴文献[7]的方法,采用一阶自相关系数 r 判断序列 $\{y_1, y_2, \dots, y_n\}$ 的自相关性。自相关函数定义为

$$r = \frac{1}{n-1} \sum_{i=1}^{n-1} y_i y_{i+1} - \left(\frac{1}{n-1} \sum_{i=1}^{n-1} y_i \right) * \left(\frac{1}{n-1} \sum_{i=1}^{n-1} y_{i+1} \right) \quad (1)$$

式中, $r \in [-1/4, 1/4)$ 。当 $r = -1/4$ 时,序列 $\{y_1, y_2, \dots, y_n\}$ 一阶负自相关;当 $r = 1/4$ 时,序列一阶正自相关;当 $r = 0$ 时,序列不相关。

本文从解释变量序列、自相关序列 2 方面进行组合预测,具体的预测流程如图 1 所示。

2 非零需求预测

2.1 需求发生原因判定

首先,需要建立一种机制来判别非零需求是由需求自相关引起的,还是由解释变量引起的。然后将由解释变量影响导致的需求和由自相关引起的需求,分别用相应的方法进行预测。

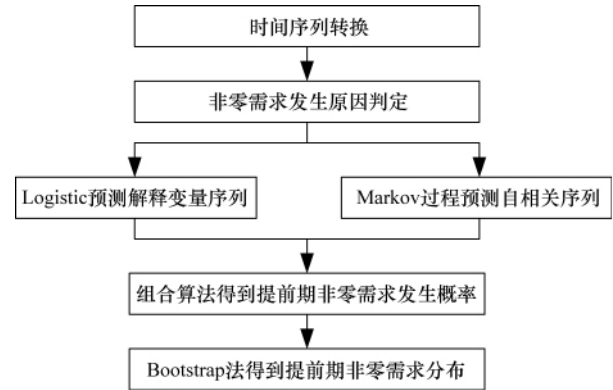


图 1 组合预测流程

Fig. 1 Flowchart of the combined forecasting

算法 1 及步骤如下:

步骤 1 设定 \bar{r} 的初始值,令 $t=1$ 。

步骤 2 当 $t \leq n$ 时,转到步骤 3;否则转到步骤 5。

步骤 3 如果 $y_t = \hat{y}_t = 1$,则令 $S_t = \{y_1, y_2, \dots, y_{t-1}, y_t\}, S'_t = \{y_1, y_2, \dots, y_{t-1}, 0\}$,计算 S_t 的相关系数 r_1 和 S'_t 的相关系数 r' ;如果 $|r_1| < \bar{r}$ 或 $|r_1| < r'_1$,令 $y_t = 0$ 。

步骤 4 令 $\bar{y}_t = y_t, t=t+1$,转到步骤 2。

步骤 5 结束。

2.2 Logistic 回归预测

采用分析分类变量常用的 Logistic 回归模型^[8]。假设有 m 个解释变量 $x_i(i=1,2,\dots,m)$,用 $\{(x'_1, x'_2, \dots, x'_m) | t=1,2,\dots,n\}$ 表示关于 $\{y_1, y_2, \dots, y_n\}$ 的解释变量的时间序列。

根据 Logistic 函数取得 Logistic 回归模型为

$$\hat{p}_t = \frac{1}{1 + \exp \left[- \left(a + \sum_{i=1}^m \beta_i x'_i \right) \right]} \quad (2)$$

式中, \hat{p}_t 为在给定序列自变量 x'_t 时的事件发生概率。利用极大似然估计,可以得到参数 α 和 β 。这样,一旦得到各个时间段的解释变量构成的样本,并同时得到其事件是否发生的观测值,就能够使用这些信息来描述和分析在特定条件下事件发生的概率。通过 Logistic 回归运算,便可以得到每一个非零需求的发生概率。

算法 2 及步骤如下。

步骤 1 计算初始值 $p_0(p_0 \in [0, 1])$,使得 $\sum_{i=1}^n (\text{sign}(\hat{p}_t - p_0) - y_t)$ 的值最小。其中

$$\text{sign}(\hat{p}_t - p_0) = \begin{cases} 1 & \hat{p}_t - p_0 \geq 0 \\ 0 & \text{其他} \end{cases}$$

步骤 2 对每一个 $\hat{p}_t(t=1,2,\dots,n)$,如果 $\hat{p}_t \geq p_0$,令 $\hat{y}_t = 1$;否则,令 $\hat{y}_t = 0$ 。

该算法对拟合时间序列概率 $\{p'_t | t=1,2,\dots,n\}$ 设置了一个下限 p_0 ,将大于 p_0 的预期值变为 1,否则转化为 0,可以得到

相应的导弹备件需求发生时间序列 $\{y_t | t=1, 2, \dots, n\}$, 这样处理的好处是使得 $\{\hat{y}_t | t=1, 2, \dots, n\}$ 与实际序列的 $\{y_t | t=1, 2, \dots, n\}$ 的匹配度最大。

2.3 Markov 自相关预测

由于 $\{\bar{y}_t | t=1, 2, \dots, n\}$ 是需求发生的自相关序列, 因此可以采用两种状态的一阶 Markov 过程进行模拟^[9]。由自相关序列 $\{\bar{y}_t | t=1, 2, \dots, n\}$ 可以得到各种状态的转移概率: 记 p_{00} 为连续零需求的概率, p_{01} 为零需求转为 1 的概率; p_{10} 为 1 转为 0 需求的概率; p_{11} 为连续非零需求的概率。状态转移概率矩阵为

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

利用状态转移概率矩阵, 以序列中最后需求 \bar{y}_n 的值为条件, 可以估计从状态 \bar{y}_n 到提前期 $l (l=1, 2, \dots, L)$ 的状态转移概率。

设 $\bar{y}_n = i (i=0, 1)$, 则 \bar{y}_{n+l} 的预测结果为 $P(\bar{y}_{n+l}=0) = p_{i0}, P(\bar{y}_{n+l}=1) = p_{i1}$ 。即

$$[P(\bar{y}_{n+l}=0), P(\bar{y}_{n+l}=1)] = [p_{i0}, p_{i1}] \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}^{l-1} \quad (3)$$

设 $p_{n+l}^A = P(\bar{y}_{n+l}=1), p_{n+l}^B$ 为第 l 期需求为非零的概率, 可以得到自相关需求序列的提前期非零需求发生概率的预测结果。

3 预测组合

3.1 提前期需求发生概率预测

前面已经通过 Logistic 回归预测和 Markov 过程预测得到解释变量序列非零需求发生概率和自相关序列非零需求发生概率。对两个预测结果, 通过组合算法 3, 可以得到组合的提前期非零需求发生概率的预测结果:

步骤 1 初始设定 r' , 令 $\tilde{y}_{n+l}=0, l \in \{1, 2, \dots, L\}$ 。

步骤 2 计算时间序列 $\{y_1, y_2, \dots, y_n\}$ 的自相关系数 1。

步骤 3 如果 $|r_1| \geq r'$, 令 $\tilde{y}_{n+l}=0$ 的概率为 p_{n+l}^A , 转到步骤 6; 否则, 令 $l=1$, 转到 Step 4。

步骤 4 如果 $l \leq L$, 计算 \hat{p}_{n+l} , 转到步骤 5; 否则, 转到步骤 6。

步骤 5 如果 $\hat{p}_{n+l} \geq p_0$, 令 $\tilde{y}_{n+l}=1$ 的概率为 \hat{p}_{n+l} ; 否则, 令 $\tilde{y}_{n+l}=1$ 的概率为 \hat{p}_{n+l} ; 令 $l=l+1$, 转到步骤 4。

步骤 6 结束。

3.2 改进灰自助法

设导弹备件需求样本为 $X = \{x(t) | t=1, 2, \dots, n\}$, 经过 Bootstrap 抽样^[10], 得到再抽样数据 $Y_b(u)$ 。下面进行 GM(1, 1) 二次数据拟合^[11]。

灰色系统理论要求 $Y_b(u) \geq 0$ 。为避免 $Y_b(u) < 0$, 可以作如下数据平移变换:

$$X_b^{(0)} = \{y_b(u) + q\} = \{x_b^{(0)}(u)\} \quad (6)$$

$u=1, 2, \dots, m; b=1, 2, \dots, p$

式中, q 是一个待定参数, 只要满足 $X_b(u) \geq 0$ 即可。如果已有 $Y_b(u) \geq 0$, 则 $q=0$ 。对 $X_b^{(0)}$ 做一次累加生成, 得一次累加生成序列向量为

$$X_b = \{x_b(u)\} = \left\{ \sum_{j=1}^u x_b^{(0)}(j) \right\} \quad (7)$$

设均值生成序列向量为

$$Z_b = \{z_b(u)\} = \{0.5x_b + 0.5x_b(u-1)\} \quad u=2, 3, \dots, m \quad (8)$$

在初始条件 $x_b(1) = x_b^{(0)}(1)$ 下, 累加生成序列的预报值为

$$\hat{x}_0(j+1) = \left[x_b^{(0)}(1) - \frac{c_1}{c_2} \right] e^{-c_1 j} + \frac{c_1}{c_2} \quad (9)$$

$j=0, 1, 2, \dots, m-1$

其中, $c_1 (c_1 \neq 0)$ 和 c_2 为 $(c_1, c_2)^T = (D^T D)^{-1} D^T (Y_b)^T, D = (-Z_b, I)^T, I = (1, 1, \dots, 1)$ 。

如果计算出 $c_1=0$, 则这个样本抽样失败, 需重新抽样。

为了提高模型精度, 对参数进行第二次拟合估计:

$$\hat{x}(j+1) = A e^{-c_1 j} + B \quad j=m-1, m \quad (10)$$

式中, $(A, B)^T = (G^T G)^{-1} G^T (X_0)^T, G = (E, I), I = (1, 1, \dots, 1)^T, E = (e^0, e^{-c_1}, e^{-2c_1}, \dots, e^{-(n-1)c_1})^T, X_0 = (\hat{x}_0(1), \hat{x}_0(2), \dots, \hat{x}_0(m))^T$ 。

根据累减生成, 就可以用仿真样本 Y_b 估计第 $w=m+1$ 次试验的数据为

$$\hat{y}_b(w) = \hat{x}_b(w) - \hat{x}_b(w-1) - q \quad w=m+1; b=1, 2, \dots, p \quad (11)$$

第 w 次试验有 p 个数据, 可构成如下序列向量

$$\hat{X}_w = \{\hat{y}_b(w)\} \quad b=1, 2, \dots, p; w=m+1 \quad (12)$$

估计 \hat{X}_w 中的 p 个数据的分布, 用 \hat{X}_w 中数据分布区逼近 X 的分布, 进而得到备件需求提前期的需求分布。

3.3 提前期需求分布计算

得到导弹备件非零需求发生的概率后, 还需要预测出非零需求的具体值。本文利用 Bootstrap 法产生足够多的预测值, 来得到提前期需求分布, 需求预测值表示为 Y_{LM} 。

算法 4 的步骤如下:

步骤 1 初始设定抽样次数 N , 令 $C=1$ 。

步骤 2 若 $C \leq N$, 令 $Y_{LM}=0$, 转到步骤 3; 否则, 转到步骤 5。

步骤 3 当 $l=1, 2, \dots, L$ 时, 运用算法 3, 得到 \tilde{y}_{n+l} ; 若 $\tilde{y}_{n+l}=1$, 在实际需求 $d_i (i=1, 2, \dots, n)$ 中以相同概率随机抽取非零需求 d'_i , 令 $\tilde{d}_{n+l}=d'_i$; 否则, 令 $\tilde{d}_{n+l}=0$ 。则 $Y_{LM} = Y_{LM} + \tilde{d}_{n+l}$ 。

步骤 4 输出 Y_{LM} , 令 $C=C+1$, 转到步骤 2。

步骤 5 结束。

4 实例分析

对某单位 2001—2010 年导弹保障情况进行数据收集, 选取 8 种关键件, 将每半年的领用量进行累积, 以此作为备件的需求量, 从而得到以半年为时间单位的备件需求时间序列。将 2001—2007 年 14 期作为训练样本, 2008—2010 年 6

期的时间序列作为检验样本。

首先以某型电子模块为例,综合导弹装备的使用和保障情况,可知除备件自身的可靠性之外,战备时间和通电时间对备件的需求影响较大。其装备使用、保障以及备件需求情

况如表 1 所示(影响因素为归一化后的值)。在组合预测方法的计算中,设定 $\bar{r}=0.08, r'=0.10$ 。在 Croston 法的计算中,设定平滑常数 $\alpha=0.25^{[5]}$ 。

表 1 保障数据统计

Table 1 Statistics of the support data

序号	战备时间	通电时间	需求量/件	序号	战备时间	通电时间	需求量/件
1	0	0.7561	2	11	0.6484	0.5972	0
2	0.1045	0.5317	1	12	0.7146	0.9413	4
3	0.2386	0.3275	0	13	0.6874	0.8475	3
4	0.5824	0.5879	3	14	0.5137	0.6728	0
5	0.3152	0.2871	0	15	0.7847	0.7249	3
6	0.2975	0.5716	1	16	0.4317	0.4017	0
7	0.6791	0.8594	4	17	0.5758	0.6137	2
8	0.4583	0.4387	0	18	1	0.9736	5
9	0.4937	0.3274	0	19	0.4129	0.4635	0
10	0.5639	1	5	20	0.8696	0.8659	4

将装备的战备时间和通电时间作为 Logistic 回归模型的解释变量。将备件需求发生转换的阈值 p_0 设定为 0.50。模型参数的估计及检验采用最大似然统计分析方法,一些专业统计软件如 SPSS 软件都提供 Logistic 回归分析功能。对 Logistic 回归方程的显著性检验采用 3 种检验方法:总体零假设(即自变量对事件发生可能性无影响作用的无效假设)的似然比检验、记分检验和 Wald 检验^[8]。由表 2 可知, χ^2 作用不显著,3 种检验方法的 P 值均小于 0.05,因此回归模型具有显著性意义。

表 2 检验总体零假设 ($\beta=0$)

Table 2 Testing global null hypothesis ($\beta=0$)

检验类型	χ^2	自由度	P 值
似然比检验	10.483	4	0.003
记分检验	8.671	4	0.004
Wald 检验	7.396	4	0.025

图 2 给出了用指数平滑法、回归分析法、Croston 法、MB 法和组合法预测的未来 6 期该型电子模块备件需求的结果。从图中可以看出,指数平滑法和回归分析法 2 种连续性预测方法的预测误差最大,超过了平均需求的 70%;组合法的预测结果相比 Croston 和 MB 2 种方法得到显著改善,仅在第 2 期和第 5 期零需求的预测上存在偏离,但预测值绝对误差仍小于 0.5。

预测需求是否发生的准确度用非零需求预测正确率 (Ratio of Occurrence Accuracy, ROA) 表示,需求量预测的准确度采用常用的平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 表示。由于 Croston 法只能预测提前期平均需求,不能预测提前期每一期的需求情况,因此无法计算 ROA。图 3 给出未来 6 期的 8 种备件预测结果,综合 8 种备件预测结果得出,组合法的 ROA 比 MB 方法提高

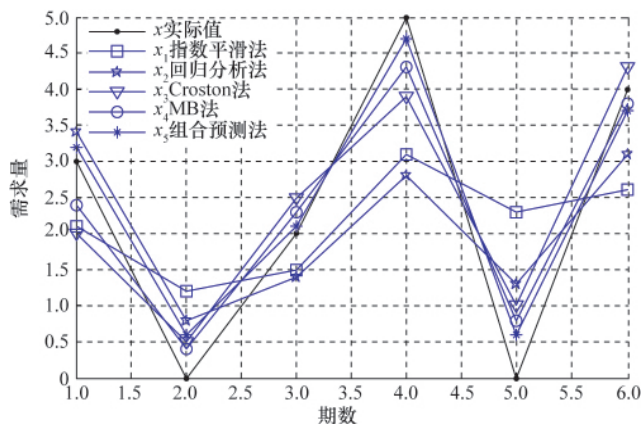


图 2 预测结果对比

Fig. 2 Comparison of forecasting results

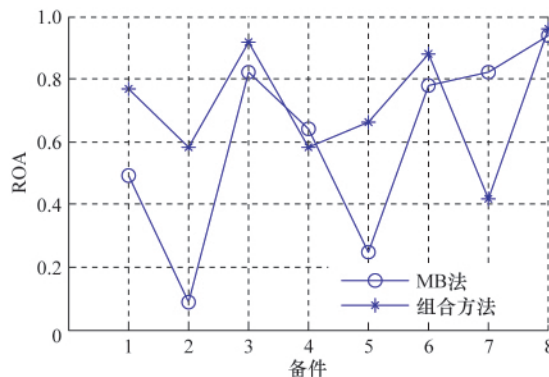


图 3 未来 6 期的 ROA 比较

Fig. 3 Comparison of ROA in the six periods

35.6%。当提前期为 2 期时,3 种方法对 7 种备件预测结果的数量准确度比较如图 4 所示。组合法对所有备件平均准确度达到 92%,远高于 Croston 法的 68.5%和 MB 法的 79.3%。

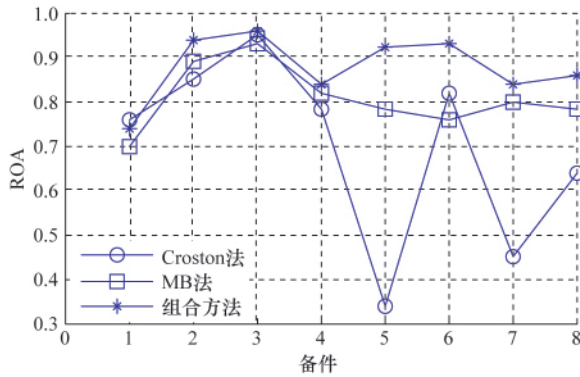


图4 未来6期的数量预测准确度比较

Fig. 4 Comparison of forecasting result in the six periods

5 结论

针对间断性导弹备件需求预测问题,提出了从解释变量序列和自相关序列两个方面进行组合预测的思路。Croston方法和 Markov Bootstrap 法的比较表明,本文提出的组合预测模型在需求是否发生的准确度和需求量精度上明显优于前述两种方法,预测结果可信度高。

参考文献 (References)

- [1] Chung S H, Kang H Y, Peam W L. A service level model for the control wafers safety inventory problem [J]. International Journal of Advanced Manufacturing Technology, 2005, 26(5): 591-597.
- [2] Leven E, Segerstedt A. Inventory control with a modified croston procedure and Erlang distribution [J]. International Journal of Production

- Economics, 2004, 90: 361-367.
- [3] Tian Y M, Shen H L, Zhang L, et al. Utility water supply forecast via a GM (1,1) weighted Markov chain [J]. Journal of Zhejiang University: Science A (Applied Physics & Engineering), 2010, 11(9): 677-682.
- [4] Zhong Y H, Zhao L, Liu Z B, et al. Using a support vector machine method to predict the development indices of very high water cut oilfields[J]. Petroleum Science, 2010, 7(3): 379-384.
- [5] Croston J D. Forecasting and stock control for intermittent demands[J]. Operational Research Quarterly, 1972, 42(3): 289-303.
- [6] Gooijer J G D, Vidiella A A. Forecasting threshold cointegrated systems [J]. International Journal of Forecasting, 2004, 20(2): 237-253.
- [7] Williams T M. Stock control with sporadic and slow-moving demand[J]. Journal of the Operational Research Society, 1984, 35(10): 939-948.
- [8] Hua Z S, Zhang B, Yang J, et al. A new approach of forecasting intermittent demand for spare parts inventories in the process industries [J]. Journal of the Operational Research Society, 2007, 58(1): 52-61.
- [9] 农吉夫. 马尔可夫链模型及其在旱涝预测的应用[J]. 广西民族大学学报:自然科学版, 2010, 16(1): 70-75.
- Nong Jifu. Markov chain model and its application in drought and flood forecasting [J]. Journal of Guangxi University for Nationalities: Natural Science Edition, 2010, 16(1): 70-75.
- [10] 金星, 洪延姬. 系统可靠性评定方法 [M]. 北京: 国防工业出版社, 2005.
- Jin Xin, Hong Yanji. Assessment methods of system reliability [M]. Beijing: National Defense Industry Press, 2005.
- [11] 刘思峰, 党耀国, 方志耕. 灰色系统理论及其应用 [M]. 北京: 科学出版社, 2004.
- Liu Sifeng, Dang Yaoguo, Fang Zhigeng. Grey system theory and its application[M]. Beijing: Science Press, 2004.

(责任编辑 齐志红)

·学术动态·



中国科协召开奋力实现中国梦首场报告会

2013年5月24日,中共中央宣传部指导、中国科协与贵州省人民政府主办的中国科协弘扬科学道德践行“三个倡导”奋力实现中国梦报告会在贵阳举行。

钱学森同志秘书、中国人民解放军总装备部研究员涂元季,中国载人航天工程总设计师、研究员周建平,暨南大学教授、中国工程院院士刘人怀,分别从钱学森同志身边工作人员角度、从承担国家重大科技项目的一线科技工作者角度、从经历了新中国科技事业发展辉煌历程的科技工作者角度,讲述我国不同时代的德学双馨的科技工作者如何通过不懈努力,为实现中华民族伟大复兴的中国梦铸就坚强基石,感召广大科技工作者和青年学生弘扬科学道德,践行“三个倡导”,积极投身实现中国梦的伟大实践,为推动科技事业健康发展、建设创新型国家做出更大贡献。

详见中国科协网 <http://www.cast.org.cn/n35081/n35096/n10225918/14729862.html>。