

# 聚类 and 马尔科夫方法结合的城市汽车行驶工况构建

姜 平 石 琴 陈无畏

合肥工业大学,合肥,230009

**摘要:**利用因子分析法提取了与汽车燃油消耗相关的2个因子(平均速度和速度标准偏差),根据这两个因子,通过两步聚类分析,把一系列汽车行驶工况分为3类,即平均速度为20km/h以下的行驶工况、20~40km/h的行驶工况和40km/h以上的行驶工况。每类行驶工况都可看作一个随时间变化的马尔科夫过程,根据新的分类准则把每类行驶工况划分为加速片段、减速片段、怠速片段和匀速片段,计算片段间的转移概率。在此基础上,提出了片段间的状态转移概率与发动机比功率相结合的候选工况选择方法,建立了评价代表性行驶工况的12个特征参数的评价准则。以合肥市典型道路为例分别构建了不同类别的行驶工况。结果表明,根据燃油消耗构建的行驶工况与试验数据具有高的相似性,符合车辆实际行驶工况。

**关键词:**行驶工况;构建;马尔科夫过程;聚类

**中图分类号:**U491.255

**文章编号:**1004—132X(2010)23—2893—05

## Driving Cycle Construction Method of City Motors Based on Clustering Method and Markov Process

Jiang Ping Shi Qin Chen Wuwei

Hefei University of Technology, Hefei, 230009

**Abstract:** Two factors (the mean speed and speed deviation) with the fuels of automobiles were extracted by the factor analysis, according to two factors a series of driving cycles were divided into 3 different clusters by the two-step clustering method. The first cluster represented driving patterns of moving vehicles under 20 km/h. The second cluster consisted mostly of data with average speeds between 20 to 40 km/h. And the last cluster included data with average speeds greater than 40 km/h. Every driving cycle was taken as a stochastic Markov process with time and the driving data for each cluster were divided into acceleration segments, deceleration segments, idling segments and cruise segments according to a new classification scheme and the transition probabilities of segments were calculated. According to this basis the selection method of candidate cycles was considered by combination the transition probabilities of segments and vehicle specific power of the engine, the evaluation criteria of 12 characteristic parameters of evaluating the representative driving cycle were established. The driving cycles for different clusters were constructed with an example of typical roads in Hefei. The results show that the driving cycles with emissions have high similarity with the experiments and accords with driving cycles of real roads.

**Key words:** driving cycle; construction; Markov process; cluster

## 0 引言

汽车行驶工况分为标准工况和非标准工况。当今世界上很多国家都以标准等形式提出了不同车型在不同应用时的标准工况,如日本的10-15工况循环、欧洲经济委员会的ECE—R15工况循环、美国联邦城市及高速公路循环(CVS—C/H)。我国汽车排放测试规程采用欧洲的排放标准,往往带来很不满意的结果,因此,需要对待开发的目标车型进行工况调查和测试,并得到合适的行驶工况。所以,开发和设计汽车、考核汽车排放污染水平都需要有相应的行驶工况。

目前,国内学者大多采用单纯的短行程和聚

类分析法对汽车行驶工况进行研究<sup>[1-3]</sup>。国外很多学者采用“微路径”以及随机过程选择方法来构建行驶工况<sup>[4-5]</sup>。本文在构建汽车行驶工况时,不同于国内传统方法。传统方法未考虑汽车燃油消耗与平均速度、发动机工作状态分布等与行驶工况有关的因素,只是笼统地对整个汽车行驶工况进行构建。本文将聚类分析法、因子分析法与马尔科夫过程相结合,按照平均速度和速度标准偏差对汽车行驶工况进行类的划分,将每类行驶工况分别看作一个马尔科夫过程,计算每类行驶工况的片段转移概率和发动机比功率,提出构建候选工况的新方法。

## 1 因子分析

因子分析的基本目的就是用少数几个因子描述许多指标或因素之间的关系,把变量表示成各

收稿日期:2010—01—29

**基金项目:**国家自然科学基金资助项目(70771036);安徽省自然科学基金资助项目(090414162);合肥工业大学基金资助项目(080205F)

因子的线性组合,即将关系比较密切的几个变量归在同一类中,每一类变量就成为一个因子,从而以较少的几个因子反映原资料的大部分信息。

发动机工作状态分布一般根据发动机比功率、汽车行驶速度、瞬时加速度决定,而且它又与汽车排放、燃油消耗密切相关,因此在构建汽车行驶工况时,应考虑发动机工作状态分布。比功率综合了速度、加速度、坡度以及风阻等参数,其物理意义为发动机输出功率与机动车质量的比值,其拟合公式如下<sup>[6]</sup>:

$$P_a = v[1.1a + 9.81\text{atan}(\sin\theta) + 0.132] + 0.000\,302v^3$$

(1)

式中, $v$ 为车辆行驶速度,m/s; $a$ 为车辆行驶瞬态加速度, $\text{m/s}^2$ ; $\theta$ 为道路坡度, $(^\circ)$ ; $P_a$ 为发动机比功率,kW/t。

参考国外发动机工作状态分布<sup>[7]</sup>,考虑国内混合交通实际情况,将我国城市典型道路行驶工况燃油效率划分为表 1 所示的分布。表 1 的工作状态分布是一个 14 维变量(B1~B14),这些变量分别由速度分布、比功率来确定。每个变量类似于一个单元,每一个单元类似于一个“小抽屉”,在这个“小抽屉”中,存放着该工况下测点的平均燃油消耗。行驶工况是逐秒采集的,每秒所观察的数据被分配在表 1 中。如 B1 表示该工况下汽车制动时平均燃油消耗量,B2 表示该工况下汽车怠速时平均燃油消耗量,B3 表示该工况下汽车比功率小于 0 且速度大于 0、小于 40km/h 时的平均燃油消耗量,依此类推,其他变量表示在该工况下不同比功率和速度区间的平均燃油消耗量。

表 1 行驶工况燃油消耗分布的划分

汽车制动(B1)		
怠速(B2)		
比功率分布 区间(kW/t)	速度分布区间(km/h)	
	(0,40]	(40,+∞)
$(-\infty,0)$	B3	B9
$[0,3)$	B4	B10
$[3,6)$	B5	B11
$[6,9)$	B6	B12
$[9,12)$	B7	B13
$[12,+\infty)$	B8	B14

本文应用统计分析软件 SPSS15.0 中的因子分析方法减少变量,采用主成分分析法提取公因子,提取特征值大于 1 的成分为主成分。因子模型的旋转方法选取方差最大变换法。对合肥市 5 条典型道路工作状态分布进行主成分分析,特征碎石图见图 1。前 2 个主成分的特征值为 1.47,累计方差贡献率为 84.3%,即前 2 个主成分包含了原有变量 84.3%的信息,所以可以取这 2 个因子来代替原有的变量。

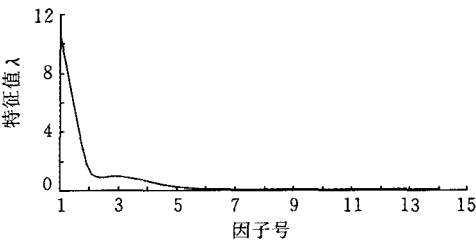


图 1 特征值碎石图

通过对因子模型进行旋转,得到旋转后的因子载荷矩阵如表 2 所示。根据旋转后的因子载荷大小来判别因子主要解释的变量,本文选取载荷大于 0.8 的变量为因子的主要解释变量。因子 1 主要解释了变量 B9、B10、B11、B12、B13、B14,由表 1 可知,这些变量主要是高速状态下的变量。低速状态变量因子载荷小,如 B0、B1、B2、B3、B4、B5、B6、B7,说明因子 1 与平均速度有关,可以解释为平均速度。因子 2 主要解释了变量 B1、B3、B4、B5,对怠速状态变量 B2 影响最小,载荷在低速时先增大,中间速度时减小,高速时先减小后增大,这表明因子 2 与车速偏离平均速度的程度有关,可以解释为速度的标准偏差。

表 2 旋转后的因子载荷矩阵

变量	因子 1	因子 2
B1	0.162	0.868
B2	0.291	0.185
B3	0.408	0.885
B4	0.343	0.916
B5	0.443	0.801
B6	0.724	0.629
B7	0.662	0.516
B8	0.666	0.674
B9	0.878	0.430
B10	0.934	0.234
B11	0.946	0.272
B12	0.903	0.329
B13	0.929	0.326
B14	0.819	0.523

2 两步聚类分析

两步聚类分析是一种新型的分层聚类方法,目前一般应用在数据挖掘与多元统计的交叉领域——模式分类中,其算法适合任何尺度的变量<sup>[8]</sup>。采用上述与燃油消耗相关的两个主要因子(平均速度和速度标准偏差)进行两步聚类分析。聚类准则采用贝叶斯信息准则 Akaike(AIC)。AIC 利用  $N-L$  平均信息量最小的原则来确定系统的阶次和估计相应参数值,其定阶准则为<sup>[9]</sup>:

$$AIC(N) = -2\ln L(\theta_{ML}) + 2N$$

(2)

式中, $N$ 为模型独立未知参数的个数; $\theta_{ML}$ 为相应参数的估计值; $L(\theta_{ML})$ 为在  $\theta_{ML}$  条件下的似然函数值。

在工程应用中,对于不同阶次  $N$ ,首先用极大似然估计法估计系统的参数值,然后计算似然函数值及其  $AIC(N)$  值,以使  $AIC(N)$  值达到最小时的阶次作为该模型的阶次。本文利用 SPSS15.0 统计软件的两步聚类模型,采用 AIC 准则对合肥市典型道路的行驶工况进行聚类分析,结果可以聚类为 3 类:第 1 类行驶工况(类 1)平均速度偏低,为  $0 \sim 20 \text{ km/h}$ ,第 2 类行驶工况(类 2)平均速度中等,为  $20 \sim 40 \text{ km/h}$ ,第 3 类行驶工况(类 3)平均速度偏高,大于  $40 \text{ km/h}$ 。

### 3 各类行驶工况的构建

#### 3.1 样本的采集

本研究中试验车为轿车,试验车尽可能跟随测试路段上的车流正常行驶。速度信号采样频率为  $1 \text{ Hz}$ 。城市道路行驶工况的研究一般是从多条道路中筛选出有代表性的典型道路进行研究。典型道路的特征是能综合反映整个城市的主要交通状况,即反映整个城市道路的道路等级(快速道、主干道、次干道及支路,以及车道数、机动车非机动车混合等),交通强度(交通流量、饱和度等),交叉口密度(路段内交叉形式和数量等)和时间(早晚交通流量变化明显程度等)。本文选取了合肥市胜利路、屯溪路、明光路等 5 条道路进行样本采集,使采集的数据能反映合肥市整体道路的交通情况,统计的汽车交通流量具有代表性。

#### 3.2 数据的分离

数据分离就是把行驶工况的速度时间曲线划分成许多反映加速度变化的模型事件,包括加速模型事件、减速模型事件、怠速模型事件和匀速模型事件,这些模型事件是构建行驶工况的基础。模型事件使用  $(a_1, a_2, \delta, n)$  准则进行定义。为了确定分类的最优值,需要考虑以下目标函数和约束条件。

目标函数  $CAP$  是一个加权函数:

$$CAP = k_1 \sum \theta_1 + k_2 \sum \theta_2 - k_3 \sum \theta_3 + k_4 \sum \theta_4 + k_5 \sum \theta_5 \quad (3)$$

式中,  $k_1, k_2, k_3, k_4, k_5$  为目标函数的权值。

为实现上述目标,要满足以下 5 个约束条件:

- ① 加速度模型事件含有连续的加速度,因此,加速标准差  $\theta_1$  应该很小;
- ② 对于任何匀速事件,最大速度和最低速度差  $\theta_2$  应该小;
- ③ 对于任何加速或减速事件,最大速度和最低速度差  $\theta_3$  应大;
- ④ 任何匀速事件的平均加速度  $\theta_4$  接近 0;
- ⑤ 因为一个单独点的特征依赖于相邻点特征,所以一个好的分类标准应该包括一小部分单秒数据  $\theta_5$ 。

通过迭代法使目标函数  $CAP$  最小,得到最优的分类准则——(5, 1, 3, 2) 分类原则<sup>[10]</sup>,即加速事件定义为单一瞬时加速度不小于  $2.22 \text{ m/s}^2$  或者连续观测加速度大于  $0.71 \text{ m/s}^2$ ,或持续 3s 或以上时间,而且要实现  $3.2 \text{ km/h}$  以上的速度递增;减速片段定义为单一瞬时加速度不大于  $-2.22 \text{ m/s}^2$ ,或任何连续观测加速度不大于  $-0.71 \text{ m/s}^2$ ,持续 3s 或以上时间,而且实现  $3.2 \text{ km/h}$  以上的速度递减。其余的被列为怠速或匀速事件。这种分类原则考虑了现实世界的行驶工况,匀速模式事件可以包括一些小的加速和减速模型事件。按照这个分类原则,行驶工况片段划分结果如图 2 所示。怠速模型事件指速度为 0 的模型事件。

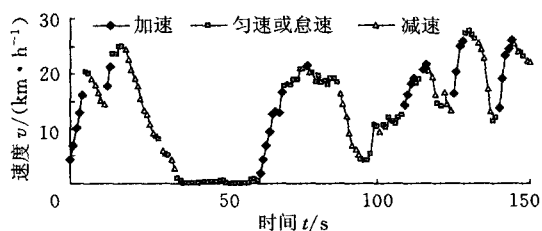


图 2 速度时间曲线片段划分结果

#### 3.3 转移矩阵的计算

马尔科夫过程是一个随机过程,记为  $Z_t, t = 1, 2, \dots, T$ ,  $Z_t$  代表每个状态的模型事件。当前时刻  $t$  状态的概率只与前一个时刻  $t-1$  状态有关。换句话说,汽车下个时间  $t+1$  的状态(如加速、减速、匀速和怠速片段)的概率只与当前时间  $t$  的加速、减速、匀速和怠速片段有关。

在马尔科夫链中,假设  $X$  是方程  $X = pX$  时( $p$  为转移矩阵,  $X$  为  $p$  的特征向量)的唯一解,则  $X$  为马尔科夫链的静态分布。对于一个静态的马尔科夫链,随着马尔科夫链的增长,整个模型事件分布保持不变。应用这个性质,根据模型事件转移矩阵构建的代表性行驶工况能代表整个试验数据的行驶工况。

为了估计每类行驶工况的转移矩阵,运用 Lee 等<sup>[11]</sup> 的最大似然估计法进行估计。对于一个固定的马尔科夫过程,应用贝叶斯公式得到连续的有序状态的概率为

$$P(Z_0, Z_1, \dots, Z_t) = P(Z_0) \prod_{i=1}^t P(Z_i | Z_{i-1}) \quad (4)$$

假定每个过程都是重复观察的,  $N_{rs}$  为  $t-1$  时刻状态  $r$  转移到  $t$  时刻状态  $s$  的事件数,则  $N$  次重复观察的公式为

$$P(Z_0, Z_1, \dots, Z_T | N) \propto P(Z_0) \prod_{r,s} P_{rs}^{N_{rs}} \quad (5)$$

根据极大似然函数,得到状态转移概率方程

$$P_n = N_n / \sum N_n \quad (6)$$

依据上述原理,得到合肥市典型道路不同类别的行驶工况的片段转移概率,如表3所示。例如,在类1工况中,若当前状态是减速片段,则下个状态为加速片段的概率是53%,为匀速片段的概率是38%,为怠速片段的概率是9%。类3工况常为高速工况,通常没有怠速,故只有3个状态。

表3 片段转移概率 %

当前状态		下个状态			
		加速	减速	匀速	怠速
类1	加速	0	50	50	0
	减速	53	0	38	9
	匀速	51	42	0	7
	怠速	57	0	43	0
类2	加速	0	53	47	0
	减速	49	0	45	6
	匀速	51	48	0	1
	怠速	84	0	16	0
类3	加速	0	46	54	
	减速	55	0	45	
	匀速	48	52	0	

### 3.4 候选工况的构建

构建行驶工况的基本思路是截取若干段实测的数据片段,拼成一个完整的速度—时间历程。随机选择一个小于100s的片段作为开始片段,即初始状态。初始状态的起始速度可以为0。下个状态的选择,即片段的选择一般根据速度、发动机比功率和转移概率来选择。依照以下步骤构建候选工况:

(1)初始状态一旦确定,它的终点速度和比功率就确定了,假设为当前状态。下个状态的初始速度与上个片段最终速度绝对值误差小于0.5km/h,而且下个状态初始发动机比功率与当前状态的终点比功率相比不能超过15kW/t。

(2)根据转移矩阵和上面提到的速度与比功率原则,选择下个状态。

(3)依据上述原则,不断选择下一个状态,直到候选工况达到理想的长度为止。

### 3.5 代表性工况的选取

为了从众多候选工况中选取合适的代表性行驶工况,本文考虑2个参数,即候选工况与试验数据的速度—加速度频率差 $d_1$ ,4种模型事件(加速、减速、怠速、匀速)时间比例差值的绝对值 $d_2$ :

$$d_1 = \sum_a \sum_v |f_{r(a,v)} - p_{(a,v)}| \quad (7)$$

$$d_2 = \sum_{m=1}^4 |f_m - p_m| \quad (8)$$

式中, $f_{r(a,v)}$ 为候选工况速度 $v$ 和加速度 $a$ 的比值; $p_{(a,v)}$ 为试验数据中速度 $v$ 和加速度为 $a$ 的比值; $f_m$ 为候选工况中第 $m$ 种模型事件的时间比值; $p_m$ 为试验数据中第 $m$ 种模型事件的时间比值。

速度—时间频率按照速度间隔5km/h、加速度间隔0.278m/s<sup>2</sup>进行计算。

综合考虑 $d_1$ 与 $d_2$ ,选择综合加权测试法(composite performance measure, CPM)进行评价,即取CPM最小的候选工况为代表性行驶工况:

$$CPM = d_1 w_1 + d_2 w_2 \quad (9)$$

式中, $w_1$ 为 $d_1$ 的标准偏差的倒数; $w_2$ 为 $d_2$ 的标准偏差的倒数。

## 4 结果分析

以合肥市5条典型道路为例,采用因子分析和两步聚类分析方法把汽车行驶工况分成3类,再分别对每类行驶工况按照马尔科夫方法构建代表性行驶工况,并与试验数据进行比较。特征参数包括平均速度 $v_m$ 、速度标准方差 $\sigma_v$ 、正加速度的平均值 $a_{v+}$ 、负加速度的平均值 $a_{v-}$ 、怠速时间比例 $p_i$ 、加速时间比例 $p_a$ 、减速时间比例 $p_d$ 、匀速时间比例 $p_c$ 、正的发动机比功率平均值 $P_+$ 、负的发动机比功率平均值 $P_-$ 、发动机比功率平均值 $P_a$ 、发动机比功率方差 $\sigma_{vsp}$ 。

每类行驶工况所构建的代表性行驶工况与试验数据的特征参数比较见表4。可以看出,低速行驶工况怠速比例高,高速行驶工况怠速比例为0。类1、类2和类3所构建的代表性行驶工况与试验数据特征参数平均误差分别为3.47%、1.26%和1.28%。表明构建的行驶工况与试验数据相符,比较合理。

表4 试验数据与代表性行驶工况的特征参数比较

特征参数	试验	类1	试验	类2	试验	类3
$v_m$ (km/h)	17.57	17.81	24.6	25.3	46.72	45.81
$\sigma_v$ (km/h)	15.42	12.46	19.45	18.95	15.27	14.72
$a_{v+}$ (m/s <sup>2</sup> )	0.68	0.63	0.65	0.62	0.49	0.46
$a_{v-}$ (m/s <sup>2</sup> )	-0.64	-0.6	-0.61	-0.59	-0.52	-0.49
$p_i$ (%)	14.74	14.75	12.85	12.5	0	0
$p_a$ (%)	30.96	29.68	31.22	31.43	36.72	36.41
$p_d$ (%)	31.85	32.35	32.96	32.82	34.32	32.91
$p_c$ (%)	22.45	23.22	22.97	23.25	28.96	30.68
$P_+$ (kW/t)	3.11	3.03	3.99	4.11	5.59	5.57
$P_-$ (kW/t)	-2.8	-2.71	-3.66	-3.72	-4.5	-4.75
$P_a$ (kW/t)	0.69	0.66	0.97	0.96	2.5	2.45
$\sigma_{vsp}$	4.1	3.81	5.15	5.1	6.4	6.29

为了检验各类行驶工况模型事件分布是否与试验数据分布一致,将3类代表性行驶工况的片段转移概率与试验数据的片段转移概率进行 $\chi^2$ 检验,3类检验概率 $P$ 值都大于0.05,说明构建行驶工况与试验数据来自同一样本。构建后的代表性行驶工况的 $\chi^2$ 检验统计值、 $d_1$ 、 $d_2$ 和CPM值如表5所示。

表 5 代表性行驶工况周期、评价参数和  $\chi^2$  分布检验

	类 1	类 2	类 3
周期 (s)	712	721	715
$d_1$	0.47	0.46	0.43
$d_2$	0.23	0.34	0.12
CPM 值	33.33	31.77	31.64
$\chi^2$ 检验统计值	6.148	3	4.39
自由度	3	3	2
概率值	0.105	0.231	0.11

类 1、类 2、类 3 三类工况构建的代表性行驶工况分别如图 3、图 4 和图 5 所示。

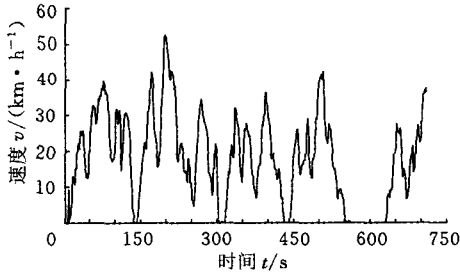


图 3 类 1 代表性行驶工况

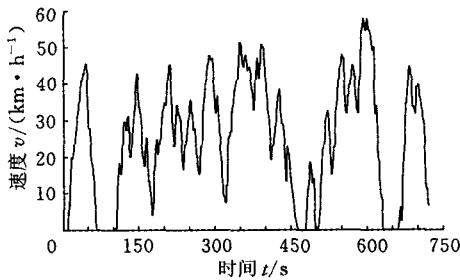


图 4 类 2 代表性行驶工况

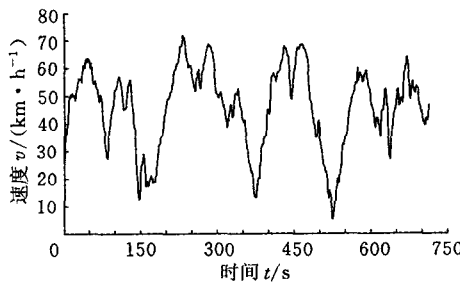


图 5 类 3 代表性行驶工况

5 结论

(1) 本文先采用两步聚类分析方法把行驶工况分成低速、中速、高速 3 类行驶工况, 然后采用马尔科夫方法构建行驶工况。构建行驶工况时考虑了与燃油消耗相关的因素(国内发动机工作状态分布), 因此构成的代表性行驶工况能较好地估计汽车的实际燃油消耗。

(2) 利用新的分类准则将行驶片段联系为特殊的模型事件来定义行驶变量, 将速度—时间曲线划分为加速、减速、匀速和怠速片段, 根据速度、比功率和转移矩阵求取各片段间的转移概率, 建

立了新的特征参数评价准则和代表性行驶工况的选取原则。

(3) 利用组合方法构建的代表性行驶工况与试验数据的特征参数相比, 平均相对误差小, 即构建的代表性工况能准确地反映城市道路上汽车的实际行驶工况。

参考文献:

[1] 吕林, 锁国涛. 不同路况下汽车行驶工况的测试与分析[J]. 武汉理工大学学报, 2006, 30(5): 781-784.

[2] 张建伟, 李孟良, 艾国和, 等. 车辆行驶工况与特征的研究[J]. 汽车工程, 2005, 27(2): 220-224.

[3] 杜青, 杨延相, 郑伟, 等. 机动车实际道路排放特性及若干影响因素的研究[J]. 内燃机学报, 2002, 20(4): 297-302.

[4] Lin J, Niemeier D A. An Exploratory Analysis Comparing a Stochastic Driving Cycle to California's Regulatory Cycle[J]. Atmospheric Environment, 2002, 36(38): 5759-5770.

[5] Eisinger D S, Niemeier D A, Stoeckenius T, et al. Collecting Driving Data to Support Mobile Source Emissions Estimation[J]. Journal of Transportation Engineering, 2006, 132(11): 845-854.

[6] 姚志良, 贺克斌, 王岐东, 等. IVE 机动车排放模型应用研究[J]. 环境科学, 2006, 27(10): 1928-1933.

[7] John K, Mitch C, Megan B, et al. MOVES2004 Software Design Reference Manual Draft[M]. Ann Arbor: Environmental Protection Agency, 2004.

[8] 陈超, 邹滢. SPASS 15.0 中文版常用功能与应用实例精简[M]. 北京: 电子工业出版社, 2009.

[9] 张济民, 王开文, 池茂儒. 基于 AIC 信息准则法的摆式列车倾摆伺服系统建模研究[J]. 中国铁道科学, 2003, 24(6): 6-9.

[10] Zhen Dai, Niemeier D A, Eisinger D. Driving Cycles: a New Cycle — building Method that Better Represents Real — world Emissions. Task Order No. 66[R]. Berkeley: Department of Civil & Environmental Engineering University of California, 2008.

[11] Lee T C, Judge G G, Zellner A. Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data [M]. Amsterdam: North-Holland Publishing Company, 1970.

(编辑 苏卫国)

作者简介: 姜 平, 女, 1974 年生。合肥工业大学交通运输工程学院副教授。主要研究方向为汽车安全与节能。发表论文 10 余篇。石 琴, 女, 1963 年生。合肥工业大学交通运输工程学院教授。陈尤畏, 男, 1951 年生。合肥工业大学机械与汽车学院教授、博士研究生导师。