第 46 卷 第 12 期　　合肥工业大学学报(自然科学版)　　Vol. 46 No. 12
2023 年 12 月　JOURNAL OF HEFEI UNIVERSITY OF TECHNOLOGY (NATURAL SCIENCE)　Dec. 2023

DOI:10.3969/j.issn.1003-5060.2023.12.001

基于短行程特征聚类的移动源排放清单构建方法

李兵兵[1]，康　宇[2,3,4]，曹　洋[2,3,4]，李亚民[4]，许镇义[2]

(1.安徽省生态环境监测中心,安徽 合肥　230071;2.合肥综合性国家科学中心人工智能研究院,安徽 合肥　230088;3.中国科学技术大学 自动化系,安徽 合肥　230026;4.中国科学技术大学 先进技术研究院,安徽 合肥　230088)

摘　要:文章基于主成分分析(principal component analysis,PCA)和 *K*-means 聚类的短行程法,构建福州市轻型车的行驶工况。首先对原始数据进行预处理、划分运动学片段并计算其特征参数,运用 PCA 对划分后片段进行参数降维,再对其进行 *K*-means 聚类分析,形成典型片段库;然后分别根据最小参数偏差法和最大相关系数法挑选片段组成行驶工况,对比后输出最终工况,并对最终行驶工况进行有效性验证;最后利用 VT-Micro(Virginia Tech microscopic) 模型并结合合成工况,计算单车排放因子,构建 2020 年福州市机动车排放清单。研究结果可为移动源排放清单构建研究提供参考。

关键词:行驶工况重构;主成分分析(PCA);*K*-means 聚类;最小参数偏差法;最大相关系数法

中图分类号:X734.2;U495　　**文献标志码**:A　　**文章编号**:1003-5060(2023)12-1585-10

Mobile source emission inventory construction based on short-stroke feature clustering

LI Bingbing[1]，KANG Yu[2,3,4]，CAO Yang[2,3,4]，LI Yamin[4]，XU Zhenyi[2]

(1. Anhui Eco-environment Monitoring Center, Hefei 230071, China; 2. Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China; 3. Department of Automation, University of Science and Technology of China, Hefei 230026, China; 4. Institute of Advanced Technology, University of Science and Technology of China, Hefei 230088, China)

Abstract: Based on the short-stroke method of principal component analysis(PCA) and *K*-means clustering, this paper constructs the driving conditions of light-duty vehicles in Fuzhou City. Firstly, the original data is preprocessed, the kinematic segments are divided and their characteristic parameters are calculated. Then, PCA is used to reduce the parameter dimension of the divided segments, and *K*-means cluster analysis is performed to form a typical segment library. According to the minimum parameter deviation method and the maximum correlation coefficient method, the segments are selected to form the driving conditions, and the final driving conditions are output after comparison, and the validity of the final driving conditions is verified. Finally, the Virginia Tech microscopic(VT-Micro) model is used to calculate the single vehicle emission factor combined with the synthetic working conditions, and the motor vehicle emission inventory in Fuzhou City in 2020 is constructed. The study can provide reference for the construction of mobile source emission inventory.

Key words: driving condition reconstruction; principal component analysis(PCA); *K*-means clustering; minimum parameter deviation method; maximum correlation coefficient method

收稿日期:2022-09-16;**修回日期**:2022-12-02

基金项目:国家自然科学基金资助项目(61725304;62033012;62103124);安徽省科技重大专项资助项目(202003a07020009)

作者简介:李兵兵(1977—),男,安徽桐城人,安徽省生态环境监测中心高级工程师;
许镇义(1993—),男,江苏宿迁人,博士,合肥综合性国家科学中心人工智能研究院副研究员,通信作者,E-mail:xuzhenyi@ustc.edu.cn.

0 引 言

近年来,随着我国城市化进程和社会经济的快速发展,机动车保有量快速增长,移动源排放会造成细颗粒物和光化学烟雾污染,已成为我国城市大气污染重要来源。排放清单作为移动源污染定量核算的重要手段,可以用于大气污染防治措施制定与溯源分析。

车辆行驶工况是针对某一类车辆,在特定区域和时间段下,描述该类车辆行驶规律的速度-时间曲线。随着车速增加,燃油车细颗粒物排放物质量浓度会明显升高,在车辆加速情况下,车辆排放的颗粒物数量、质量浓度也明显增高[1]。通过研究道路移动源行驶工况来估计车辆排放,对移动源排放清单构建具有重要参考价值。

美国和欧洲开展汽车排放标准研究较早,已建立符合其交通特点与行驶状态的行驶工况,如新标欧洲循环测试(New European Driving Cycle,NEDC)工况[2]、美国联邦试验规程75(Federal Test Procedure 75,FTP-75)工况[3]。NEDC工况测试时间较短,速度分布比较规则,行驶里程较短且加速度比较稳定,其循环时长为1 180 s,行驶里程为11.007 km,平均速度为33.6 km/h,主要在欧洲、中国、澳大利亚等国家和地区使用。FTP-75工况是美国环境保护署用于排放标准制定的测试工况,其由车辆冷启动阶段、车辆热稳定阶段和车辆热启动阶段3部分组成,工况总时长为1 874 s,行驶里程为17.77 km,平均速度为34.1 km/h,主要使用的国家和地区包括美国、加拿大、南美洲等。

随着城市路网规模不断扩大,车辆高速行驶时间不断增加,欧美和日本排放实验领域专家共同制定了全球统一的轻型车测试工况(Worldwide harmonized Light-duty vehicle Test Cycle,WLTC)[4]。WLTC工况主要包括低速阶段、中速阶段、高速阶段和额外高速阶段,工况运行时长为1 800 s,运行距离为23.3 km,平均速度为46.3 km/h,其充分考虑车辆性能差异,对不同功率质量比的车辆采用不同测试循环,能够客观真实地反映车辆的实际排放状况。

当前行驶工况合成方法主要包括基于马尔科夫决策的工况构建方法和基于聚类算法的短行程构建方法。

文献[5]基于马尔科夫决策方法,使用特定的选取原则进行片段选取,形成聚类片段库,从而从物理意义上反映车辆的行驶规律;该方法适用范围广泛,不限道路类型,但在起始和终止片段的选取方面缺乏统一的选取原则。而起始和终止片段的选取,特别是起始片段的选取,将直接影响所构建工况模型的精确性。

文献[6]首次提出基于聚类算法的短行程构建方法,该方法可以有效地与主成分分析(principal component analysis,PCA)[7]和K-means[8]等聚类方法相结合,自动选取行程片段,因此受到更多的关注。

在国内,文献[9]针对不同测试循环工况下的轻型车污染物排放特性进行研究。

由于我国汽车工业起步较晚,目前在工况测试排放标准研究方面大多沿用欧洲的NEDC工况,但是该工况与我国目前的人员驾驶习惯及车辆标准有一定出入。因此,根据我国实际道路驾驶数据,构建符合当地实际路况的行驶工况具有现实意义,对于车辆排放物及燃料消耗等测试与评价有重要参考作用。

同时,现有的移动源排放清单研究大多基于文献[10]中的排放因子本地化修正方法,采用欧洲环境署开发的宏观排放模型COPERT(Computer Program to calculate Emissions from Road Transport model)[11]、美国环境保护署应用于汽车排放源计算的IVE模型(International Vehicle Emissions Model)[12]和美国环境保护署开发的移动车辆排放模型(MOtor Vehicle Emission Simulator,MOVES)[13]等排放因子模型方法,初步探讨了全国[14]、区域[15-17]、省(直辖市)[18-21]、省会城市[22-23]等多个尺度的机动车污染物排放结构组成、时空分布特征及相应减排情景对策等。

本文采用基于PCA和K-means聚类的短行程法,构建福州市轻型车的行驶工况。首先对原始数据进行预处理、划分运动学片段并计算其特征参数,运用PCA对划分后片段进行参数降维,再对其进行K-means聚类分析形成典型片段库;然后分别根据最小参数偏差法和最大相关系数法挑选片段组成行驶工况,对比后输出最终工况,并对最终行驶工况进行有效性验证;最后利用VT-Micro(Virginia Tech microscopic)模型并结合合成工况,计算单车排放因子,构建2020年福州市机动车排放清单。

1 基于短行程特征工况的清单构建

本文提出一种基于短行程特征聚类的移动源

行驶工况排放清单构建方法，具体步骤如下：

1）对获取的车辆 GPS 轨迹数据进行预处理。

2）对轨迹的速度数据进行片段划分得到运动学片段，并根据运动学片段特征计算车辆运动学特征参数和总数据的总体分布参数。

3）利用 PCA 对运动学片段进行运动学特征参数降维。

4）使用 K-means 聚类对输入片段进行分类，得到多种类型的片段库。

5）分别按照最大相关系数法和最小参数偏差法从各类片段库中挑选输入片段，构建合成行驶工况，并对比 2 种方法所构建工况的综合指标。

6）引入 VT-Micro 模型，计算合成工况排放清单。

1.1　数据采集

本文采用未规定路线的自主行驶法采集福州市出租车行驶数据。不限定出车时间及地点，将 GPS 定位的终端设备安装于实验车辆上，进行数据远程传输和存储。为使得采集的轻型车行驶数据可以代表福州市乘用车行驶特点，行驶路线涵盖福州市主要道路。

福州市区分为 5 个行政区，下辖 5 个县及 2 个县级市。5 个行政区分别为鼓楼区、台江区、晋安区、仓山区及马尾区，前 3 个为福州市主城区。5 个县包括闽侯县、闽清县、永泰县、罗源县及连江县，2 个县级市为福清市、长乐市。福州市区道路网呈“棋盘式＋环路”分布。

考虑到大部分车辆行驶轨迹都集中于二环以内，且大部分的医院、商场、学校等也设于二环内，本文选取的主要实验区域为鼓楼区、台江区、晋安区及仓山区；考虑到福州大学、闽江学院及马尾区也会占有城市部分交通量，选择二环快速、三环快速、福州绕城高速及主城区内的主干道作为实验主要线路。

利用车载自动诊断系统（On-Board Diagnostics，OBD）接口采集发动机运行状态信息，车辆运行参数包括轨迹点采样时间、GPS 车速、X 轴加速度、Y 轴加速度、Z 轴加速度、轨迹采样点经纬度位置坐标、发动机转速、扭矩百分比、瞬时油耗、油门踏板开度、空燃比、发动机负荷百分比和进气流量，累计采集 331 550 条 GPS 轨迹数据，采样时间间隔为 1 s，车辆运行参数的 4 种统计数据见表 1 所列。

表 1　车辆运行参数 4 种统计数据

参数	最小值	最大值	均值	标准差
GPS 车速/(km/h)	0	116.600	25.383	23.369
X 轴加速度/(m/s^2)	0	1.026	−0.003	0.048
Y 轴加速度/(m/s^2)	−1.188	0.180	−0.420	0.062
Z 轴加速度/(m/s^2)	−1.440	−0.252	−0.875	0.048
经度/(°)	0	119.683	1 19.251	3.613
纬度/(°)	0	26.489	25.908	0.822
发动机转速/(r/min)	125.000	4 450.000	1 174.675	428.445
扭矩百分比/%	0	94.000	25.463	10.979
瞬时油耗/(mL/s)	0	69.280	13.332	23.659
油门踏板开度	0	0.470	0.045	0.054
空燃比	0	18.6	14.4	1.6
发动机负荷百分比/%	9.000	100.000	31.184	12.488
进气流量/(m^3/min)	0.740	64.420	5.386	4.124

1.2　数据预处理

需要对采集的原始 GPS 轨迹和车辆信息数据进行预处理，以去除无效片段及各类异常值，保证用于构建工况的行驶数据有效可靠。

考虑到轨迹数据采样时间间隔为 1 s，车辆运行前后时刻速度不会突变过大，因此可以根据前后时刻的速度来填充当前时刻速度，若前后时刻速度全部缺失，直接删除该数据。速度缺失值处理策略见表 2 所列，插值处理结果如图 1 所示。

表 2　速度缺失值处理策略

前一时刻速度 v_{t-1}	下一时刻速度 v_{t+1}	当前时刻速度 v_t^*
未缺失	未缺失	$v_t^*=(v_{t-1}+v_{t+1})/2$
未缺失	缺失值	$v_t^*=v_{t-1}$
缺失值	未缺失	$v_t^*=v_{t+1}$
缺失值	缺失值	删除该数据

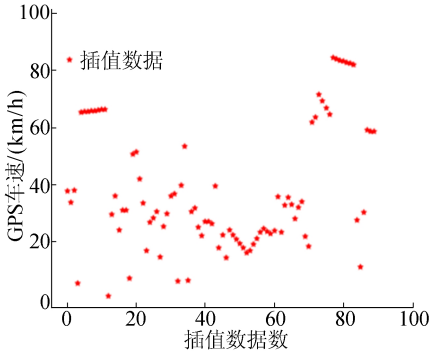


图 1 GPS 车速缺失记录插值处理结果

经车载设备直接采集的原始数据会包含部分异常数据。由于高层建筑或隧道遮挡,部分时刻 GPS 信号丢失,使得采集数据中的时间不连续,对于此类数据直接进行丢弃处理。若车辆长时间低速且间断行驶(速度小于 10 km/h),则可视为怠速状态,默认最长怠速时间为 180 s,多于 180 s 则按异常值进行丢弃。

1.3 运动学片段划分与特征参数计算

由于行驶过程中车辆速度随时间的变化特点不同,可以根据加速度和速度将不同类型的速度-时间片段定义为不同的运动状态,便于后续分析和聚类。本文定义以下 4 种运动状态:

1) 怠速状态。车速值为 0 但发动机转速值不为 0 的运动状态。

2) 加速状态。车辆加速度大于 0.15 m/s^2 的运动状态。

3) 减速状态。车辆加速度小于 −0.15 m/s^2 的运动状态。

4) 匀速状态。车辆加速度在 −0.15 ~ 0.15 m/s^2 之间且车速值不为 0 的运动状态。

将运动学片段定义为前一个停车时刻(速度值降为 0)到下一个停车时刻之间的速度-时间曲线,因此运动学片段会由 1 段速度值为 0 的怠速片段和 1 段速度值不为 0 的行驶片段组成。

由于从上一停车时刻到下一停车时刻,速度必须经过上升和下降过程,车辆必然会出现加速、减速、匀速等状态,运动学片段一般会出现怠速、加速、减速、匀速 4 种运动状态。典型运动学片段组成如图 2 所示。不同的运动学片段是车辆在不同地段不同时间段的行驶数据,其速度-时间曲线的时间长度及形状特点存在差异,而每个运动学片段都由若干个速度-时间点构成,通过定义各类特征值来区分不同曲线,也能定量刻画不同曲线之间的差异性,为后续的聚类分析提供依据。

同时,在构建最终行驶工况后,还需要一些指标来定量分析和评价其有效性和准确性,以及与车辆真实行驶情况的贴近性。

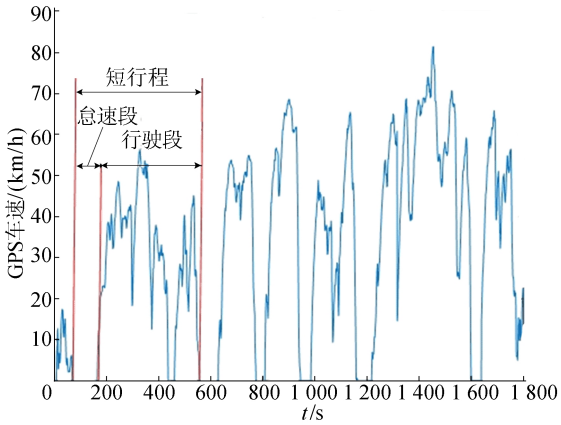


图 2 车辆运动学片段组成

因此,选取的特征参数必须尽可能全面、准确地描述运动学片段的运动特点。根据参数用途,将定义的特征参数分为 2 个大类:① 车辆运动特征参数,主要用来描述运动学片段的运动特征,可作为后续聚类分析的分类依据;② 数据总体分布特征参数,主要用于最终构建工况的检验与有效性验证。选取的 15 个车辆运动特征参数和 10 个数据总体分布特征参数分别见表 3、表 4 所列。

表 3 车辆运动特征参数及其含义

参数	含义	参数	含义
t_a/s	加速时间	v_m/(km/h)	平均速度
t_d/s	减速时间	v_{mr}/(km/h)	平均行驶速度(怠速状态除外)
t_e/s	匀速时间	a_{max}/(m/s^2)	最大加速度
t_c/s	怠速时间	a_{min}/(m/s^2)	最大减速度
T/s	片段时长	a_a/(m/s^2)	平均加速度
S/m	行程距离	a_d/(m/s^2)	平均减速度
v_{max}/(km/h)	最大速度	a_{sd}/(m/s^2)	加速度标准差
v_{sd}/(km/h)	速度标准差		

表 4 数据总体分布特征参数及其含义

参数	含义	备注
p_a/%	加速时间比例	
p_d/%	减速时间比例	
p_e/%	匀速时间比例	
p_c/%	怠速时间比例	
$p_{[0,20)}$/%	[0,20) km/h 速度段比例	低速
$p_{[20,40)}$/%	[20,40) km/h 速度段比例	中速
$p_{[40,60)}$/%	[40,60) km/h 速度段比例	中速
$p_{[60,80)}$/%	[60,80) km/h 速度段比例	高速
$p_{[80,100)}$/%	[80,100) km/h 速度段比例	高速
$p_{[100,120)}$/%	[100,120) km/h 速度段比例	高速

首先对加速度特征参数进行计算,在采样数

据中，陀螺仪加速度是无法直接使用的，需要根据速度数据单独计算车辆加速度，即

$$a_{t+1}=\frac{1\ 000}{3\ 600}\frac{v_{t+1}-v_t}{\Delta t} \tag{1}$$

其中：Δt 为采样时间间隔，$\Delta t=1$ s；v_{t+1}、v_t 分别为 $t+1$、t 时刻速度；a_{t+1} 为 t 时刻到 $t+1$ 时刻的加速度。

其次计算各类行程时长特征，计算公式为：

$$T=N \tag{2}$$

$$t_{\mathrm{a}}=N_{\mathrm{a}} \tag{3}$$

$$t_{\mathrm{d}}=N_{\mathrm{d}} \tag{4}$$

$$t_{\mathrm{c}}=N_{\mathrm{c}} \tag{5}$$

$$t_{\mathrm{e}}=N-N_{\mathrm{a}}-N_{\mathrm{d}}-N_{\mathrm{c}} \tag{6}$$

其中：N 为运动学片段的总点数，考虑数据采样时间间隔为 1 s，采样点数等于运动时长；N_{a} 为加速度大于 0.15 m/s^2 的总点数；N_{d} 为加速度小于 −0.15 m/s^2 的总点数；N_{c} 为速度为 0 的总点数。

然后依次计算 S、$v_{\max}$、v_{m}、v_{mr}、v_{sd}、$a_{\max}$、$a_{\min}$、a_{a}、a_{d} 及 a_{sd}，计算公式分别为：

$$S=\sum_{t=1}^{N}\frac{v_t}{3.6} \tag{7}$$

$$v_{\max}=\max\{v_1,v_2,\cdots,v_N\} \tag{8}$$

$$v_{\mathrm{m}}=3.6\frac{S}{T},\ v_{\mathrm{mr}}=3.6\frac{S}{T-t_{\mathrm{c}}} \tag{9}$$

$$v_{\mathrm{sd}}=\sqrt{\frac{1}{N-1}\sum_{t=1}^{N}(v_t-v_{\mathrm{m}})^2} \tag{10}$$

$$a_{\max}=\max\{a_1,a_2,\cdots,a_N\} \tag{11}$$

$$a_{\min}=\min\{a_1,a_2,\cdots,a_N\} \tag{12}$$

$$a_{\mathrm{a}}=\frac{\mathrm{sum}\{a_t\mid a_t>0.15\ \mathrm{m/s^2},t=1,2,\cdots,N\}}{t_{\mathrm{a}}} \tag{13}$$

$$a_{\mathrm{d}}=\frac{\mathrm{sum}\{a_t\mid a_t<-0.15\ \mathrm{m/s^2},t=1,2,\cdots,N\}}{t_{\mathrm{d}}} \tag{14}$$

$$a_{\mathrm{sd}}=\sqrt{\frac{1}{N-1}\sum_{t=1}^{N}a_t^2} \tag{15}$$

将所有采样样本按照式(1)～式(15)进行参数计算，得到 2 818 个运动学片段的车辆运动特征参数和样本的总体分布特征参数。

4 种运动状态时间占比分别为：

加速，25.02%；减速，20.35%；

匀速，30.43%；怠速，24.19%。

速度分布占比分别为：

$p_{[0,20)}$，43.35%；$p_{[20,40)}$，26.43%；

$p_{[40,60)}$，18.99%；$p_{[60,80)}$，6.49%；

$p_{[80,100)}$，4.10%；$p_{[100,120)}$，0.60%。

由运动状态分布结果可知，福州市轻型车的加速、减速、匀速和怠速 4 类状态时间分布较为均匀，其中匀速时间段比例最大（占比 30.4%），由此可见福州市轻型车的行驶较平稳，匀速驾驶状态占比较大。而根据速度分布结果可知，大部分时间车辆速度处于[0，40) km/h 的中低速区间，高速([60，120) km/h)占比小（占比 11.2%），具有较强的区域特征。因此，构建特定区域或时间段行驶工况对于车辆行驶特征研究是必要的。

在聚类后，会生成 K 个（K 为聚类类别数）不同的典型片段库，在进行片段挑选构成最终工况时，需要计算能代表典型片段库的综合参数，以反映该片段库的车辆行驶特点。基于一定挑选原则，根据每个片段车辆运动学参数和典型片段库整体参数的接近程度，来挑选构成最终工况的运动学片段，采用相关系数法挑选片段时需要片段综合参数。本文选取 p_{a}、p_{d}、p_{c}、p_{e}、$v_{\max}$、v_{m}、v_{mr}、$a_{\max}$、$a_{\min}$、$p_{[0,20)}$、$p_{[20,40)}$、$p_{[40,60)}$、$p_{[60,80)}$ 共 13 个指标作为片段库的综合参数。考虑到速度段[80，120) km/h 的数据占比不足 8%，纳入综合指标会导致较大偏差，因此，保证[0，80) km/h 区间的速度分布贴合，即可保证构建工况的准确性。将典型片段库中每种状态时间总和除以典型片段库的总运行时间得到各类状态占比。

1.4　基于 PCA 和 K-means 的典型工况构建

短行程法将不同运动学片段根据特征参数值相似度分为几类，形成能反映不同交通特点和行驶状况的典型运动学工况片段库，再根据一定的原则从库中挑选出片段组合，形成最终的工况曲线。本文先通过 PCA 将冗余参数去除，只留下几个主要特征组合，再通过运动学片段聚类构建典型工况片段。

利用 PCA 对运动学片段进行运动学特征参数降维，样本数据包含 m 个个体、n 个评价指标，则相应指标矩阵为：

$$\boldsymbol{X}=[\boldsymbol{X}_1\quad \boldsymbol{X}_2\quad \cdots\quad \boldsymbol{X}_n]=\begin{pmatrix} x_{11} & \cdots & x_{1n}\\ \vdots & & \vdots\\ x_{m1} & \cdots & x_{mn}\end{pmatrix} \tag{16}$$

重新组合成的新综合指标为：

$$\begin{cases}\boldsymbol{F}_1=a_{11}\boldsymbol{X}_1+a_{12}\boldsymbol{X}_2+\cdots+a_{1n}\boldsymbol{X}_n,\\ \boldsymbol{F}_2=a_{21}\boldsymbol{X}_1+a_{22}\boldsymbol{X}_2+\cdots+a_{2n}\boldsymbol{X}_n,\\ \qquad\vdots\\ \boldsymbol{F}_n=a_{n1}\boldsymbol{X}_1+a_{n2}\boldsymbol{X}_2+\cdots+a_{mn}\boldsymbol{X}_n\end{cases} \tag{17}$$

在新的综合指标中,$\boldsymbol{F}_i$、$\boldsymbol{F}_j$ 为不相关特征向量。方差达到最大并要求其依次递减,满足的限制条件为:

$$a_{i1}^2+a_{i2}^2+\cdots+a_{in}^2=1 \tag{18}$$

在进行PCA之前,需对各类指标进行标准化处理,消除单位影响。标准化计算公式为:

$$\bar{a}_j=\frac{1}{n}\sum_{i=1}^{n}a_{ij} \tag{19}$$

$$\mathrm{Var}\ a_j=\frac{1}{n-1}\sum_{i=1}^{n}(a_{ij}-\bar{a}_j)^2 \tag{20}$$

$$x_{ij}=\frac{a_{ij}-\bar{a}_j}{\sqrt{\mathrm{Var}\ a_j}} \tag{21}$$

则标准化后的矩阵为:

$$\boldsymbol{X}=[\boldsymbol{X}_1\quad \boldsymbol{X}_2\quad \cdots\quad \boldsymbol{X}_p] \tag{22}$$

相关系数矩阵为:

$$\boldsymbol{R}=\begin{pmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & & \vdots \\ r_{p1} & \cdots & r_{pp} \end{pmatrix} \tag{23}$$

其中,r_{ij} 为 $\boldsymbol{X}_i$ 与 $\boldsymbol{X}_j$ 的相关系数。r_{ij} 计算公式为:

$$r_{ij}=\frac{\mathrm{Cov}(\boldsymbol{X}_i,\boldsymbol{X}_j)}{\sqrt{\mathrm{Var}\ \boldsymbol{X}_i\ \mathrm{Var}\ \boldsymbol{X}_j}} \tag{24}$$

$$\mathrm{Cov}(\boldsymbol{X}_i,\boldsymbol{X}_j)=\frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_j) \tag{25}$$

其中:$\bar{x}_i=\frac{1}{n}\sum_{j=1}^{n}x_{ij}$;$\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}$。

计算出相关系数矩阵的特征值 λ 及特征向量 $\boldsymbol{E}$,则有:

$$|\boldsymbol{R}-\lambda\boldsymbol{E}|=0 \tag{26}$$

主成分贡献占比 f_i、主成分累计贡献率 a_k 计算公式分别为:

$$f_i=\lambda_i\Big(\sum_{i=1}^{p}\lambda_i\Big)^{-1},\quad a_k=\sum_{i=1}^{k}f_i \tag{27}$$

当 $a_k>85\%$ 时,主成分可以代表绝大部分信息,主成分值可计算为:

$$\boldsymbol{Z}=[\boldsymbol{X}\boldsymbol{u}_1\quad \boldsymbol{X}\boldsymbol{u}_2\quad \cdots\quad \boldsymbol{X}\boldsymbol{u}_k]=[\boldsymbol{Z}_1\quad \boldsymbol{Z}_2\quad \cdots\quad \boldsymbol{Z}_k] \tag{28}$$

其中,$\boldsymbol{u}_k$ 为对应的主成分特征向量。

1.5 基于合成工况的单车排放因子估算

单车排放因子(emission factor,EF)是机动车行驶单位距离所排放污染物的质量,特定平均速度下排放因子 $F_{v,j}^{\mathrm{E}}$ 计算公式为:

$$F_{v,j}^{\mathrm{E}}=3\ 600\Big(\frac{1}{v}\sum R_{i,j}P_{i,v}\Big) \tag{29}$$

其中:v 为车辆的平均速度 v_{m};$R_{v,j}$ 为车辆在 v_{m} 下的第 j 种污染物排放因子;$R_{i,j}$ 为第 i 个比功率区间的第 j 种污染物排放率;$P_{i,v}$ 为 v_{m} 在第 i 个比功率区间的分布值。

由于无法直接获取机动车污染物排放率数据,需要利用其瞬时速度和加速度数据进行估算。VT-Micro 模型能够基于车辆的瞬时速度和加速度计算车辆瞬时排放,其结构简单,被广泛用于交通研究中[24]。利用 VT-Micro 模型进行交通排放测算,计算公式为:

$$R_{v,a}=0.001\exp\Big(\sum_{i=0}^{3}\sum_{j=0}^{3}\alpha_{i,j}v^ia^j/3.6\Big) \tag{30}$$

其中:v 为车辆的瞬时速度;a 为车辆的瞬时加速度;$R_{v,a}$ 为车辆在 v、a 下的排放率;$\alpha_{i,j}$ 为 v^j 和 a^j 的回归系数,根据美国橡树岭国家实验室收集的车辆排放数据[24]得到的回归系数取值见表5所列。

表5 回归系数 $\alpha_{i,j}$ 取值

系数	油耗	CO	HC	NO_x
$\alpha_{0,0}$	−0.679 439	0.887 447	−0.728 042	−1.067 682
$\alpha_{0,1}$	0.135 273	0.148 841	0.012 211	0.254 363
$\alpha_{0,2}$	0.015 946	0.030 550	0.023 371	0.008 866
$\alpha_{0,3}$	−0.001 189	−0.001 348	−0.000 093	−0.000 951
$\alpha_{1,0}$	0.029 665	0.070 994	0.024 950	0.046 423
$\alpha_{1,1}$	0.004 808	0.003 87	0.010 145	0.015 482
$\alpha_{1,2}$	−2.053 5E−5	9.322 8E−5	−0.000 103	−0.000 131
$\alpha_{1,3}$	5.540 9E−8	−7.06E−7	6.18E−7	3.28E−7
$\alpha_{2,0}$	−0.000 276	−0.000 786	−0.000 205	−0.000 173
$\alpha_{2,1}$	8.332 9E−5	−0.000 926	−0.000 549	0.002 876
$\alpha_{2,2}$	9.37E−7	4.918 1E−5	3.759 2E−5	−5.866E−5
$\alpha_{2,3}$	−2.479 6E−8	−3.14E−7	−2.13E−7	2.4E−7
$\alpha_{3,0}$	1.487E−6	4.616E−6	1.949E−6	5.69E−7
$\alpha_{3,1}$	−6.132 1E−5	4.614 4E−5	−0.000 113	−0.000 321
$\alpha_{3,2}$	3.04E−7	−1.41E−6	3.31E−6	1.943E−6
$\alpha_{3,3}$	−4.672 34E−9	8.172 4E−9	−1.739E−8	−1.257 4E−8

根据 1.3 节构建的低速、中速、高速 3 种典型合成工况，结合式(29)分别计算 3 种工况下平均单车排放因子 $\bar{F}_K^{\mathrm{E}}$，$K\in\{$低速，中速，高速$\}$，最后根据 3 种工况时间占比 T_K 得到最终平均合成工况排放因子 $\bar{F}^{\mathrm{E}}$，即

$$\bar{F}^{\mathrm{E}}=\sum_{K}(\bar{F}_K^{\mathrm{E}}T_K) \tag{31}$$

2　实验和讨论

对划分的运动学片段采用 PCA 进行分析，得到前 4 个主成分 $M_1\sim M_4$，其累计贡献率为 88.64%，大于 85%，则前 4 个主成分可以代表样本的大部分信息。主成分方差及贡献率见表 6 所列，前 4 个主要成分的载荷矩阵见表 7 所列。

表 6　车辆运动特征参数主成分方差、贡献率计算结果

主成分	方差	贡献率/%	累计贡献率/%
M_1	6.532	43.67	43.67
M_2	3.253	22.44	66.11
M_3	2.367	13.16	79.27
M_4	1.626	9.37	88.64
M_5	0.987	5.26	93.90
M_6	0.351	1.81	94.71
M_7	0.113	0.56	95.27

表 7　前 4 个主成分特征载荷矩阵

变量	M_1	M_2	M_3	M_4
p_{e}	0.911	0.201	0.843	0.341
p_{c}	0.271	0.957	0.038	0.412
p_{d}	0.902	0.007	0.278	−0.147
p_{a}	0.856	0.259	0.211	0.264
v_{mr}	0.874	0.392	0.189	0.423
v_{m}	0.253	−0.328	0.932	−0.177
$v_{\max}$	0.912	0.159	0.552	0.102
v_{sd}	−0.253	0.433	0.818	0.231
$a_{\min}$	0.412	0.058	0.227	0.752
$a_{\max}$	−0.101	0.374	−0.098	0.828
a_{a}	0.231	0.326	0.074	0.769
a_{d}	0.387	0.536	0.391	0.254
a_{sd}	0.412	0.426	0.763	0.136

对划分的运动学片段进行 K-means 聚类，得到典型片段库，并根据戴维森-堡丁指数(Davies-Bouldin index，DBI)选择最优聚类数。DBI 通过计算距离来度量相同类之间的相似性、不同类之间的差异性，计算公式为：

$$I_{\mathrm{DB}}=\frac{1}{k'}\sum_{i=1}^{k'}\max\frac{\mathrm{avg}\ C_i+\mathrm{avg}\ C_j}{d_{\mathrm{cen}}(u_i,u_j)},\quad j\neq 1 \tag{32}$$

其中：k'为簇别数；avg C_i 为第 i 类样本点到簇中心 u_i 的欧式距离平均值；$d_{\mathrm{cen}}(u_i,u_j)$为第 i 类簇中心 u_i 与第 j 类簇中心 u_j 之间的欧式距离。DBI 越小，表示相同类运动学片段之间的相似性越高、不同类运动学片段之间的差异性越大，聚类效果越好。选取 DBI 值第 1 次出现局部最小值时的 k'值，作为最优聚类数 k_0。根据最小参数偏差法和最大相关系数法，分别对各类片段进行排序挑选，计算各类片段库占总体片段库的时间比，确定各类片段库中选取的片段时长，根据排序优先级挑选片段形成行驶工况。最终选择聚类数目 $k'=3$，聚类效果通过 t 分布随机邻近嵌入(t-distributed stochastic neighbor embedding，t-SNE)方法可视化，如图 3 所示。

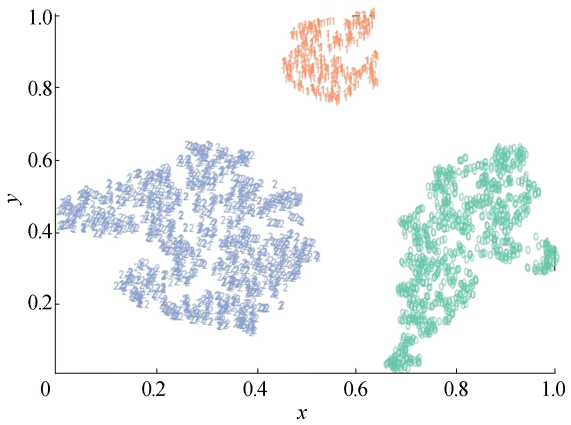


图 3　运动学片段聚类 t-SNE 降维可视化

进一步分析各类库所代表的车辆行驶特征，根据参数整体分布情况，可将运动学片段分为低速(类别 2)、中速(类别 0)、高速(类别 1)3 类，计算典型运动学片段库综合参数，结果见表 8 所列。

表 8　3 类典型片段库综合参数

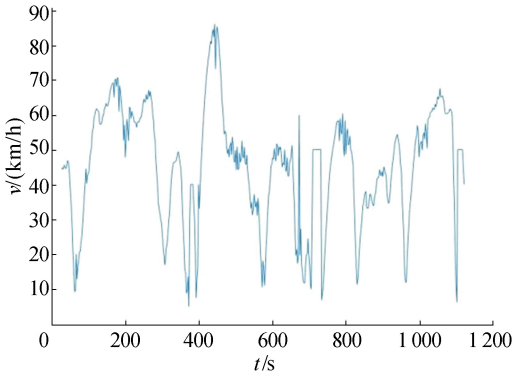
参数	类别 0（中速）	类别 1（高速）	类别 2（低速）
累计时间/s	63 424	35 673	84 516
p_{e}/%	26.50	43.30	18.80
p_{c}/%	25.50	7.60	46.00
p_{d}/%	21.10	21.70	17.00
p_{a}/%	26.90	27.40	18.20
v_{mr}/(km/h)	30.610	51.830	15.510
v_{m}/(km/h)	22.820	47.880	8.370
$v_{\max}$/(km/h)	79.800	117.100	45.900
v_{sd}/(km/h)	18.122	26.592	10.305
$p_{[0,20)}$/%	43.60	16.00	82.74
$p_{[20,40)}$/%	36.20	20.80	17.10
$p_{[40,60)}$/%	18.80	32.00	0.20
$p_{[60,80)}$/%	1.30	17.40	0
$p_{[80,100)}$/%	1.00	12.00	0
$p_{[100,120)}$/%	0	1.70	0
$a_{\min}$/(m/s^2)	0.600	0.673	0.432
$a_{\max}$/(m/s^2)	−8.000	−8.920	−6.050
a_{sd}/(m/s^2)	9.520	14.800	7.800

类别 1 运动学片段匀速比例大,速度最大值高达 117.100 km/h,且中高速([40,80) km/h)及高速([80,120) km/h)占比最大,可定义为高速行驶片段,且加减速时间均匀、匀速比例大,路况畅通,属于高速公路或快速道行驶。

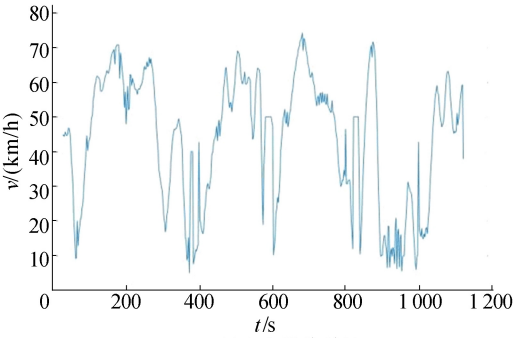
类别 2 运动学片段低速占比(82.74%)非常大,怠速比例大,平均速度仅 8.370 km/h,可定义为低速行驶片段,属于市内拥堵路段或低速次干道行驶。

类别 0 运动学片段以中低速([0,40) km/h)为主,4 种运动状态时间均匀,平均行驶速度适中,最大速度介于类别 1 与类别 2 之间,为中速行驶类型,属于主干道行驶。低速行驶时长(84 516 s)占比最大,中速时长(63 424 s)占比次之,高速时长(35 673 s)占比最少,可见福州市乘用车整体以低速拥堵行驶情况为主,主干道行驶情况较少,高速公路及快速道行驶情况极少。

所有运动学片段被分为高速、中速、低速 3 类典型片段库。为客观反映车辆真实的驾驶特性,最终的工况应该涵盖这 3 类典型运动学片段,并且在最终工况中的各类片段时间比例与整体数据中比例一致。本文采用总体最大相关系数法(方法 1)与最小参数偏差法(方法 2)2 种方法进行片段挑选,3 类典型片段 2 种方法选择结果如图 4～图 6 所示。

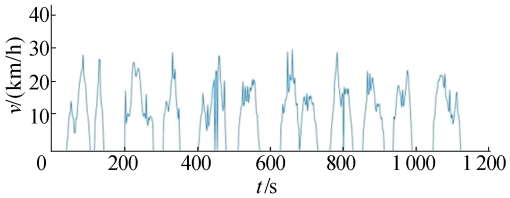


(a) 最大相关系数法

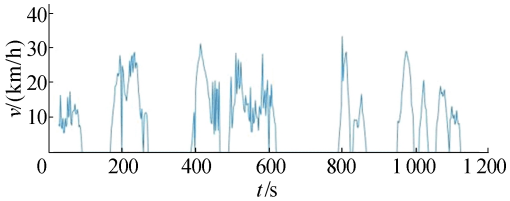


(b) 最小参数偏差法

图 4 高速片段 2 种方法选择结果

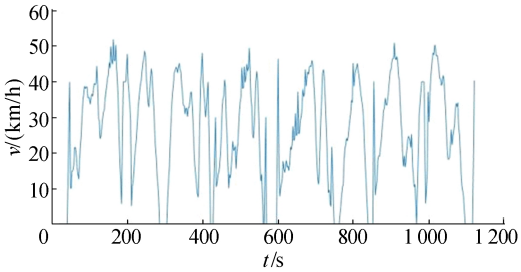


(a) 最大相关系数法

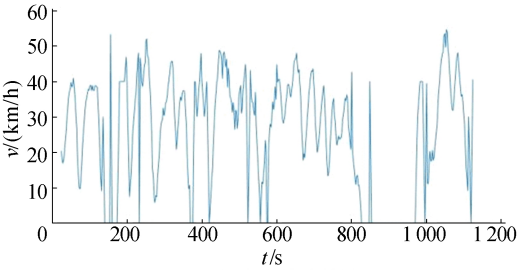


(b) 最小参数偏差法

图 5 低速片段 2 种方法选择结果



(a) 最大相关系数法



(b) 最小参数偏差法

图 6 中速片段 2 种方法片段选择结果

对于 2 种片段选择方法构建出的工况,用平均相对误差来检验工况的有效性。平均相对误差即所有参数相对误差的平均值。

2 种片段选择方法所构建工况的相对误差见表 9 所列。

由表 9 可知:最大相关系数法(方法 1)和最小参数偏差法(方法 2)的平均相对误差分别为 7.49%、10.88%,最大相关系数法的平均相对误差更小,其整体工况更贴近于真实驾驶情况;在运动状态占比和速度占比方面,最大相关系数法明显更加贴近总体数据。因此,选择最大相关系数法输出工况为最终工况。

最终构建的工况参数在整体上和真实行驶情况很贴近,尤其是在速度段[0,40) km/h 比例与 4 种运动状态比例方面吻合度极高,在速度段[60,80) km/h和速度标准差方面吻合度欠佳。

表 9　2 种方法片段合成相对误差

参数	取值			相对误差/%	
	总体	方法 1	方法 2	方法 1	方法 2
p_e/%	30.43	27.59	32.87	9.34	8.08
p_c/%	24.19	23.06	20.23	4.65	16.37
p_d/%	25.02	26.19	26.07	4.69	4.06
p_a/%	20.35	22.15	22.38	8.84	11.02
v_{mr}/(km/h)	36.943 0	34.039 2	31.831 8	7.86	13.83
v_m/(km/h)	28.007 0	26.188 5	26.664 8	6.49	4.79
v_{sd}/(km/s)	28.007 0	21.816 4	20.743 3	12.14	18.20
$p_{[0,20)}$/%	43.35	43.52	39.78	0.36	8.24
$p_{[20,40)}$/%	26.43	26.11	31.64	1.24	21.24
$p_{[40,60)}$/%	18.99	21.23	17.74	11.70	6.56
$p_{[60,80)}$/%	6.49	8.13	7.83	25.20	20.06
a_a/(m/s^2)	0.530 3	0.571 7	0.565 9	7.08	5.16
a_d/(m/s^2)	−0.657 8	−0.598 9	−0.698 7	9.01	6.15
a_{ad}/(m/s^2)	0.596 5	0.561 8	0.640 1	5.83	8.47

根据式(31)计算出单车合成工况排放因子，见表 10 所列。

表 10　单车合成工况排放因子计算结果

工况	排放因子/(g/km)			时间占比/%
	CO	HC	NO_x	
高速段	5.040 4	3.622 7E−7	2.155 8E−14	34.92
中速段	0.272 7	0.056 8	0.036 8	62.09
低速段	0.402 7	0.191 8	0.207 8	82.74
合成工况	1.258 9	0.107 9	0.108 4	

根据文献[25]，2020 年福州市汽车保有量为 1 498 695 辆，公路通车里程合计 11 616.816 km，得到 2020 年福州市机动车污染物排放清单，见表 11所列。

表 11　2020 年福州市机动车年均排放清单　　单位：t

区域	CO	HC	NO_x
福州市	21 918	1 878.6	1 886.7

3　结　　论

本文提出一种基于短行程特征聚类的移动源行驶工况排放清单构建方法，采用基于 PCA 和 K-means 聚类的短行程法，构建福州市轻型车的行驶工况。首先对原始数据进行预处理、划分运动学片段并计算其特征参数，运用 PCA 对划分后片段进行参数降维，再对其进行 K-means 聚类分析，形成典型片段库；然后分别根据最小参数偏差法和最大相关系数法挑选片段组成行驶工况，对比后输出最终工况，并对最终行驶工况进行有效性验证；最后利用 VT-Micro 模型并结合合成工况，计算单车排放因子，构建 2020 年福州市机动车排放清单。主要结论如下：

1）相比于传统的随机选择法，本文采用最大相关系数法和最小参数偏差法，分别以参数向量整体相似度最高和参数数值总体偏差最小为标准进行最终工况的片段挑选。

2）对于选定参数作出较大调整，主要是以整体参数为准，去掉 $p_{[100,120)}$（[100，120）km/h 速度段占比不足 1%，会引入较大误差）、v_{max}、a_{min} 等参数，目的是追求整体吻合度更高，大部分数据贴合真实情况，而不是追求极值或者少部分数据拟合效果。

3）在实验线路及区域选取时，根据车流量分布特点，按一定比例选取主城区与郊区实验车辆；并采用“不选定路线的自主行驶法”采集行驶数据，驾驶员可不受限制地自主驾驶，采集数据更具真实性。

本文虽然考虑行驶工况信息进行排放清单构建，但采用的 VT-Micro 排放因子模型仍然是一种基于统计回归的数学估算模型，最终的排放清单核算与实际情况仍然存在较大偏差。在未来的研究中，可考虑利用便携式排放测试系统（portable emissions measurement system，PEMS）获取的实际道路排放监测数据，对现有排放因子估算模型进行修正，以获得较准确的排放清单核算。

[参　考　文　献]

[1]　许镇义. 基于时空数据挖掘的城市移动源污染综合监管问题研究[D]. 合肥：中国科学技术大学，2020.

[2]　SHIM B J，PARK K S，KOO J M，et al. Work and speed based engine operation condition analysis for new European

driving cycle(NEDC)[J]. Journal of Mechanical Science and Technology,2014,28(2):755-761.

[3] WI H,PARK J. Analyzing uncertainty in evaluation of vehicle fuel economy using FTP-75[J]. International Journal of Automotive Technology,2013,14:471-477.

[4] TUTUIANU M,MAROTTA A,STEVEN H,et al. Development of a Worldwide harmonized Light duty driving Test Cycle (WLTC): GRPE-68-03[R/OL]. (2013-12-01)[2022-09-10]. https://unece.org/fileadmin/DAM/trans/doc/2014/wp29grpe/GRPE-68-03e.pdf.

[5] BÖHME M,PHAM V T,ROYCHOUDHURY A. Coverage-based greybox fuzzing as Markov chain[J]. IEEE Transactions on Software Engineering, 2019, 45(5): 489-506.

[6] HOU J,GAO H J,LI X L. DSets-DBSCAN:a parameter-free clustering algorithm[J]. IEEE Transactions on Image Processing,2016,25(7):3182-3193.

[7] ZHAO Z Z,SHKOLNISKY Y,SINGER A. Fast steerable principal component analysis[J]. IEEE Transactions on Computational Imaging,2016,2(1):1-12.

[8] QIN J H,FU W M,GAO H J,et al. Distributed *k*-means algorithm and fuzzy *c*-means algorithm for sensor networks based on multiagent consensus theory[J]. IEEE Transactions on Cybernetics,2016,47(3):772-783.

[9] 张鹏宇,张新宇,郝利君,等.不同测试循环工况下的轻型车污染物排放特性分析[J].车辆与动力技术,2016(4):47-49.

[10] 生态环境部.道路机动车大气污染物排放清单编制技术指南(试行)[EB/OL].(2021-05-03)[2022-09-10]. https://www.mee.gov.cn/gkml/hbb/bgg/201501/W020150107594587831090.pdf.

[11] XU Z Y,KANG Y,CAO Y,et al. Deep amended COPERT model for regional vehicle emission prediction[J]. Science China Information Sciences,2021,64:139202.

[12] 王虎,李孟良,乔维高.一种汽车源排放模型:IVE模型简介[J].中国环境监测,2007,23(5):78-81.

[13] United States Environmental Protection Agency. Motor Vehicle Emission Simulator(MOVES) 2010 user guide [EB/OL].(2010-01-01)[2022-09-10]. https://www.nrc.gov/docs/ML1127/ML112710375.pdf.

[14] 宋翔宇,谢绍东.中国机动车排放清单的建立[J].环境科学,2006,27(6):1041-1045.

[15] YANG W,YU C Y,YUAN W,et al. High-resolution vehicle emission inventory and emission control policy scenario analysis,a case in the Beijing-Tianjin-Hebei (BTH) region,China[J]. Journal of Cleaner Production,2018,203:530-539.

[16] JIANG P Y,ZHONG X,LI L Y. On-road vehicle emission inventory and its spatio-temporal variations in North China Plain[J]. Environmental Pollution,2020,267:115639.

[17] 刘永红,姚达文,黄建彰.珠三角地区机动车排放清单建立与来源分析[J].环境科学与技术,2015,38(增刊1):458-463.

[18] 李荔,张洁,赵秋月,等.基于COPERT模型的江苏省机动车时空排放特征与分担率[J].环境科学,2018,39(9):3976-3986.

[19] 李媛,石嘉诚,陈军辉,等.2010～2017年四川省机动车污染物排放趋势分析[J].环境科学,2021,42(2):643-652.

[20] SUN S D,JIANG W,GAO W D. Vehicle emission trends and spatial distribution in Shandong province,China,from 2000 to 2014[J]. Atmospheric Environment,2016,147:190-199.

[21] WANG H K,CHEN C H,HUANG C,et al. On-road vehicle emission inventory and its uncertainty analysis for Shanghai,China[J]. Science of the Total Environment,2008,398:60-67.

[22] GONG M M,YIN S S,GU X K,et al. Refined 2013-based vehicle emission inventory and its spatial and temporal characteristics in Zhengzhou,China[J]. Science of the Total Environment,2017,599/600:1149-1159.

[23] 周子航,邓也,吴柯颖,等.成都市道路移动源排放清单与空间分布特征[J].环境科学学报,2018,38(1):79-91.

[24] 蒋阳升,刘梦,王思琛,等.基于网联车轨迹重构的交通油耗和排放估计方法[J].安全与环境学报,2022,22(4):2147-2155.

[25] 福州市统计局,国家统计局福州调查队.福州市统计年鉴:2021[M].北京:中国统计出版社,2022:103-104.

(责任编辑 张淑艳)