

# 基于Markov链与SARIMAX模型的混合动力车辆速度预测

李孟岩, 陈 征, 傅梦亿

宁波工程学院理学院, 浙江 宁波

收稿日期: 2022年9月26日; 录用日期: 2022年10月24日; 发布日期: 2022年10月31日

## 摘 要

针对车辆速度预测较为复杂的问题, 提出将马尔可夫模型与SARIMAX模型引入车辆速度预测方法之中。基于美国密歇根州安阿伯市的车辆数据, 将汽车运行的速度数据处理为时间序列数据之后建立模型进行预测。详细对比分析了马尔可夫模型与SARIMAX模型在四种不同工况(低速、中速、中高速、高速)与不同预测长度(5个、10个、15个、20个)下预测车速的准确程度。经过研究可以得出: 在低速工况下, 利用SARIMAX模型和马尔可夫模型对车速进行短期预测或长期预测都是不错的选择; 在高速工况下, 长期预测应选择马尔可夫模型, 短期预测两个模型都能获得较好的效果; 在中速和中高速工况下, 长期预测应选择SARIMAX模型, 中速工况下的车速短期预测选择SARIMAX模型更合适, 中高速工况下的车速短期预测, 两种模型的预测效果欠佳。

## 关键词

马尔科夫链, SARIMAX模型, 预测对比

# Speed Prediction of Hybrid Electric Vehicle Based on Markov Chain and SARIMAX Model

Mengyan Li, Zheng Chen, Mengyi Fu

College of Science, Ningbo University of Technology, Ningbo Zhejiang

Received: Sep. 26<sup>th</sup>, 2022; accepted: Oct. 24<sup>th</sup>, 2022; published: Oct. 31<sup>st</sup>, 2022

## Abstract

Aiming at the complicated problem of vehicle speed prediction, Markov model and SARIMAX model are introduced to predict the vehicle velocity. Based on vehicle data of Ann Arbor, Michigan, USA,

this paper processes the vehicle velocity data into time series data and then substitutes them into two models. The accuracy of Markov model and SARIMAX model in predicting vehicle speed under four different working conditions (low speed, medium speed, medium high speed and high speed) and different prediction lengths (5, 10, 15 and 20) is compared and analyzed in detail. Through practical research, it can be concluded that under high-speed conditions, Markov model is selected for long-term prediction, and SARIMAX model and Markov model for short-term prediction can achieve better results; SARIMAX model is selected for long-term prediction under medium speed and medium high speed conditions. SARIMAX model is more suitable for short-term prediction of vehicle speed under medium speed conditions. For short-term prediction of vehicle speed under medium and high speed conditions, the prediction effects of the two models need to be improved.

## Keywords

Markov Chain, SARIMAX Model, Forecast Comparison

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近几年我国汽车人均保有量不断攀升, 尾气污染问题受到大众的普遍关注。在国家对环境保护日益重视的大背景下, 减少汽车尾气排放、减少车辆能源消耗成为个人和企业关注的重点, 混合动力汽车成为这一时期的关键产物[1]。混合动力汽车在行驶过程中尾气排放增加的主要阶段有: 路口红灯后起步阶段、道路拥堵时的走走停停或在车辆较少的路面的突然加速等等。因为混合动力汽车的能耗极大地依赖于车辆的速度, 需要根据各道路的实际工况来计算及预测最适合该路况的车辆速度, 从而提高整车的燃油经济性, 减少汽车尾气排放量, 实现对车辆能量管理策略的优化[2]。

对道路车速的预测方法, 目前应用较多的是基于规则的能量管理策略, 但对不同工况的适应性较差, 无法得到最优控制效果; 动态规划算法虽能得到较好的控制效果, 但须知晓全局的工况, 因此只能应用于交通仿真。许多学者运用马尔可夫模型与 ARIMA 模型解决车速预测问题, 解少博等人提出了一种基于马尔可夫链的并联 PHEB 预测型能量管理策略, 实现了档位选择和功率分配的协同优化[3]; 袁焕涛等人进行了基于马尔可夫的混合动力汽车行驶车速预测研究, 利用马尔可夫模型获得了较好的效果[4]; 张良力等人利用 ARIMA 模型对交叉路口车速进行预测, 从而实现车辆碰撞风险评估, 指出使用更多的历史数据可以提高车速预测和风险评估的精度[5]; Guo 等人利用 ARIMA 模型对车辆循环工况进行道路坡度和车速预测, 利用动态规划构建能量管理策略[6]。

目前利用马尔可夫模型与 ARIMA 模型对车速预测的对比分析研究较少, 本文引入马尔可夫模型与 SARIMAX 模型对车辆速度进行预测, 并且针对不同工况与不同预测长度对二者预测的精确度进行了对比和分析, 进而提出不同工况与不同预测长度下的最优模型选择。

## 2. 数据预处理

本文所用数据来自美国密歇根大学的研究团队, 该团队收集了美国密歇根州安阿伯市 2017 年 11 月至 2018 年 11 月, 由 264 辆汽油车、92 辆混合动力车和 27 辆混合动力电动汽车组成的多元化车队在真实世界中的行驶数据, 包含车辆的 GPS 轨迹及其燃料、能源、速度和辅助电力使用的时间序列数据, 总

里程约 374,000 英里[7]。

将每辆车从每段路开始行驶的时间戳记为 0，车辆完全驶离该路段的时间戳记为 T，速度记为 V，得到该车在该路段全程的瞬时速度变化。随机选取四种不同工况下具有代表性的车辆行驶数据，提取车辆编号、路段编号、时间戳等信息，具体见表 1。

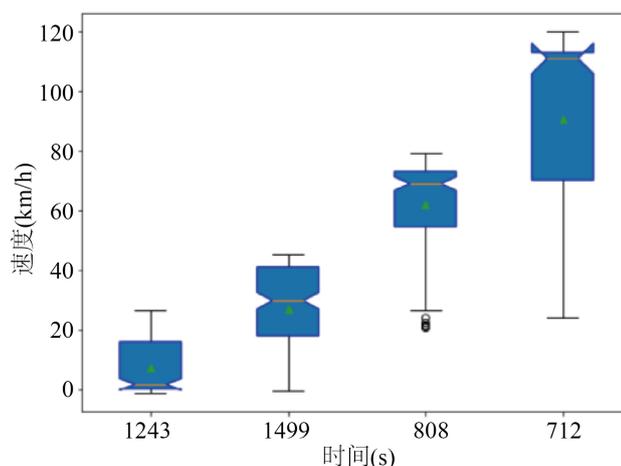
**Table 1.** Data preprocessing

**表 1.** 数据预处理

车辆编号	路段编号	时间戳(s)	数据个数	速度均值(km/h)
140	1243	0.0~183.1	178	7.34
155	1499	0.0~161.2	165	26.60
351	808	0.0~146.0	178	62.03
147	712	0.0~141.3	147	90.46

绘制这四种工况的箱线图如图 1，其中编号为 140 的车辆在 1243 路段上行驶的速度均值为 7.34 km/h，为低速工况；155 号车在 1499 路段上行驶的速度均值为 26.60 km/h，为中速工况；351 号车在 808 路段上行驶的速度均值为 62.03 km/h，为中高速工况；147 号车在 712 路段上行驶的速度均值为 90.46 km/h，为高速工况。

从图 1 的速度箱线图中可以直观地感受到四种工况的区别，且 155 号车在 1499 路段行驶的速度分布较为均匀，140 号车在 1243 路段行驶的速度呈右偏分布，351 号车在 808 路段行驶的速度和 147 号车在 712 路段行驶的速度呈左偏分布。



**Figure 1.** Speed box diagram under four different working conditions

**图 1.** 四种不同工况的速度箱线图

由于初始记录的车辆速度数据并非规律的时间序列数据，不符合 SARIMAX 模型与马尔可夫模型对于实验数据的要求，这就导致初始数据无法直接代入模型进行预测。

为了更好地满足数据要求、拟合模型进行预测，就需要对收集到的初始数据进行预处理。由于三次样条插值具有计算方便、稳定性高、光滑性好且易于操作的优良特征，于是利用 Python 工具对数据进行

了三次样条插值拟合,得到了以一秒为间隔的170个速度时间序列数据,作为标准数据进行建模,后文所提原数据均为在此经过三次样条插值拟合处理后的时间序列数据,拟合结果如图2所示,横轴为时间戳,单位为秒(s);纵轴为时间戳对应汽车瞬时速度,单位为千米每小时(km/h)。

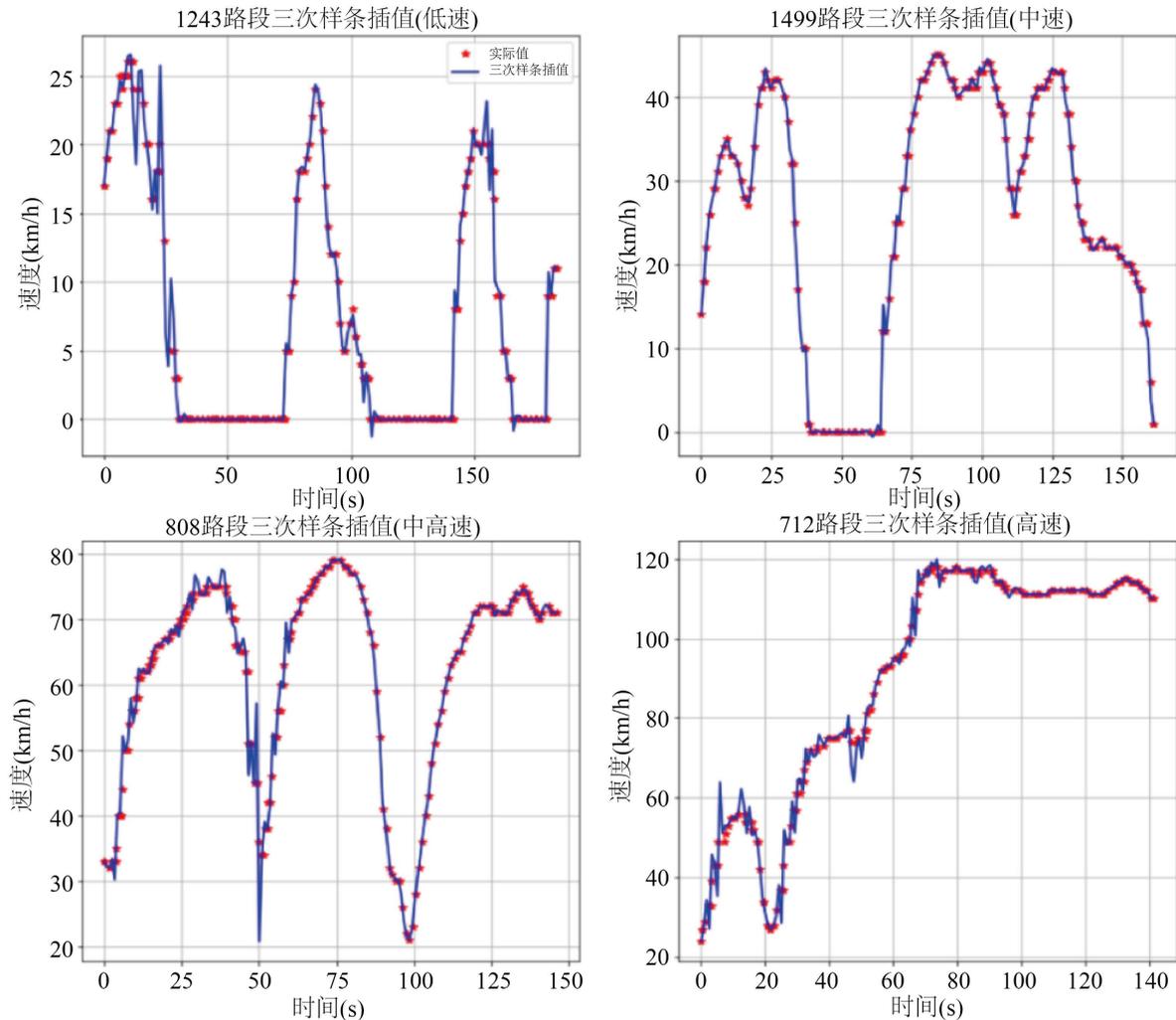


Figure 2. Cubic spline interpolation speed under four working conditions

图2. 四种工况下的三次样条插值速度

### 3. 马尔可夫预测模型

#### 3.1. 马尔科夫链

马尔可夫链是一组具有马尔可夫性质的离散随机变量的集合[8]。

对概率空间内以一维可数集为指数集的随机变量集合,若随机变量的取值都在可数集内,且随机变量的条件概率满足如下关系则被称为马尔可夫链,可数集被称为状态空间,马尔可夫链在状态空间内的取值称为状态。

设有随机过程  $X_T = \{X_t, t \in T\}$ , 其中时间  $T = \{0, 1, 2, \dots\}$ , 状态空间为  $I = \{0, 1, 2, \dots\}$ , 若对任意的正整数  $k$ , 任意  $t_i \in T$ ,  $t_i < t_{i+1}$  ( $i = 0, 1, 2, \dots, k+1$ ) 及任意非负整数  $i_0, i_1, \dots, i_{k+1}$ , 有

$$P(X_{t_{k+1}} = i_{k+1} | X_{t_0} = i_0, X_{t_1} = i_1, \dots, X_{t_k} = i_k) = P(X_{t_{k+1}} = i_{k+1} | X_{t_k} = i_k) \quad (1)$$

马尔可夫链的马尔可夫性质也被称为“无记忆性”，即  $t+1$  步的随机变量在给定第  $t$  步随机变量后与其余的随机变量条件独立[9]。隐马尔可夫模型的要素：隐含状态  $S$ 、可观测状态、初始状态概率矩阵  $\pi$ 、状态转移概率矩阵  $A$ 。

### 3.2. 建立模型

设  $X_{1t} = \{X_{1n}, n=0,1,2,\dots\}$  表示 140 号车在 1243 路段的速度，状态空间为  $I_1 = \{A, B, C, D\}$ ；  
 $X_{2t} = \{X_{2n}, n=0,1,2,\dots\}$  表示 155 号车在 1499 路段的速度，状态空间为  $I_2 = \{A, B, C, D, E\}$ ；  
 $X_{3t} = \{X_{3n}, n=0,1,2,\dots\}$  表示 351 号车在 808 路段的速度，状态空间为  $I_3 = \{A, B, C, D, E, F, G, H, I\}$ ；  
 $X_{4t} = \{X_{4n}, n=0,1,2,\dots\}$  表示 147 号车在 712 路段的速度，状态空间为  $I_4 = \{A, B, C, D, E, F, G, H, I, J, K, L, M, N, O\}$ 。

根据各工况下的数据，将速度等距划分为不同的状态，得到表 2。

**Table 2.** State division section (km/h)

**表 2.** 状态划分区间(km/h)

状态	A	B	C	D		
1243 路段(低速)	[0, 6.63)	[6.63, 13.26)	[13.26, 19.89)	[19.89, 26.52)		
状态	A	B	C	D	E	
1499 路段(中速)	[0, 5, 9.02)	[9.02, 18.04)	[18.04, 27.06)	[27.06, 36.08)	[36.08, 45.1)	
状态	A	B	C	D	.....	I
808 路段(中高速)	[20.75, 27.25)	[27.25, 33.75)	[33.75, 40.25)	[40.25, 46.75)	.....	[72.25, 79.25)
状态	A	B	C	D	.....	O
712 路段(高速)	[24, 30.39)	[30.39, 36.78)	[36.78, 43.17)	[43.17, 49.56)	.....	[113.46, 119.85)

以编号为 140 的车辆在 1243 路段下的行驶数据为例，将三次样条插值后得到的 170 个速度数据由低至高划分为 4 个状态区间，即 A, B, C, D，结果如表 3 所示。

**Table 3.** Sample of speed data change of vehicle 140 on section 1243 (km/h)

**表 3.** 140 号车在 1243 路段速度数据变动情况样本(km/h)

序号	1	2	3	4	5	6	7	8	9	10
速度	17	19.003	20.953	20.969	22.915	22.998	24.640	24.152	25.001	26.419
状态	C	C	D	D	D	D	D	D	D	D
序号	11	12	13	14	15	16	17	18	19	.....
速度	26.513	21.877	18.530	25.283	25.381	21.713	20.053	18.328	15.233	.....
状态	D	D	C	D	D	D	D	C	C	.....

同样以编号为 140 的车辆在 1243 路段下的行驶数据为例，计算一步转移概率矩阵如表 4 所示。

**Table 4.** Probability transition matrix  
**表 4.** 概率转移矩阵

	A	B	C	D
A	0.9529	0.0471	0.0	0.0
B	0.25	0.625	0.125	0.0
C	0.0435	0.0435	0.5652	0.3478
D	0.0	0.04	0.28	0.68

将三次样条插值后得到的 170 个数据划分为训练集和测试集,用前 150 个数据作为训练集建立模型,后 20 个数据作为测试集,并分别预测后 5、10、15、20 个速度状态,测试模型的预测精度。利用马尔可夫模型预测 140 号车在 1243 路段的速度状态结果如表 5 所示。

**Table 5.** Display of forecast result process  
**表 5.** 预测结果过程展示

编号	A	B	C	D	预测状态
101	0.9529	0.0471	0.0	0.0	A
102	0.9199	0.0743	0.0059	0.0	A
103	0.8954	0.0899	0.0126	0.0020	A
104	0.8763	0.0990	0.0189	0.0058	A
.....	.....	.....	.....	.....	.....

### 3.3. 预测结果与实际结果对比

低速工况的预测结果和实际结果对比如表 6 所示,当预测 15 个速度数据时,准确率为 100%;当预测 20 个数据时,准确率为 80%。当车辆低速行驶时,中短期的马尔可夫模型预测较为准确。

**Table 6.** Comparison between predicted results and actual results of Section 1243 (low speed)  
**表 6.** 1243 路段(低速)预测结果与实际结果对比

编号	151	152	153	154	155	156	157	158	159	160
实际状态	A	A	A	A	A	A	A	A	A	A
预测状态	A	A	A	A	A	A	A	A	A	A
编号	161	162	163	164	165	166	167	168	169	170
实际状态	A	A	A	A	A	A	B	B	B	B
预测状态	A	A	A	A	A	A	A	A	A	A

中速工况的预测结果和实际结果对比如表 7 所示,当预测 5 个速度数据时,准确率为 80%;预测 10 个速度数据时,准确率为 40%;预测 15 个速度数据时,准确率为 26.67%;预测 20 个速度数据时,准确率为 20%。因此在中速工况下,马尔可夫模型在短期之内的预测准确率较高,对长期的速度预测结果不太理想。

**Table 7.** Comparison between predicted results and actual results of section 1499 (medium speed)**表 7.** 1499 路段(中速)预测结果与实际结果对比

编号	151	152	153	154	155	156	157	158	159	160
实际状态	C	C	C	C	C	C	C	C	C	C
预测状态	C	C	C	C	D	D	D	D	D	D
编号	161	162	163	164	165	166	167	168	169	170
实际状态	C	C	C	B	B	B	B	B	A	A
预测状态	D	D	D	D	D	E	E	E	E	E

中高速工况的预测结果和实际对比结果如表 8 所示, 当预测 5 个速度数据时, 准确率为 40%; 预测 10 个速度数据时, 准确率为 30%; 预测 15 个速度数据时, 准确率为 53.33%; 预测 20 个速度数据时, 准确率为 40%。可见在中高速工况下, 马尔可夫模型的预测效果并不理想。

**Table 8.** Comparison between predicted results and actual results of 808 section (medium and high speed)**表 8.** 808 路段(中高速)预测结果与实际结果对比

编号	151	152	153	154	155	156	157	158	159	160
实际状态	H	H	I	I	I	I	I	I	I	H
预测状态	H	H	H	H	H	H	H	H	H	H
编号	161	162	163	164	165	166	167	168	169	170
实际状态	H	H	H	H	H	H	H	H	H	H
预测状态	H	H	H	H	H	I	I	I	I	I

高速工况的预测结果和实际对比结果如表 9 所示, 当预测 5 个速度数据时, 准确率为 100%; 当预测 10 个速度数据时, 准确率为 50%; 当预测 15 个速度数据时, 准确率为 40%; 当预测 20 个速度数据时, 准确率为 55%。在高速工况下, 短期内的马尔可夫模型预测效果较好。

**Table 9.** Comparison between the predicted results and the actual results of section 712 (expressway)**表 9.** 712 路段(高速)预测结果与实际结果对比

编号	151	152	153	154	155	156	157	158	159	160
实际状态	N	N	N	N	N	O	O	O	O	O
预测状态	N	N	N	N	N	N	N	N	N	N
编号	161	162	163	164	165	166	167	168	169	170
实际状态	O	O	O	O	N	N	N	N	N	N
预测状态	N	N	N	N	N	N	N	N	N	N

## 4. SARIMAX 预测模型

### 4.1. SARIMAX 模型

用于汽车速度预测的时间序列数据是指在不同时间点上所收集到的该汽车的瞬时速度数据, 用于车辆速度随时间变化的情况, 反映了车辆速度随时间的变化的状态。在实际操作过程当中我们通常会选取有限的观测数据进行实验。

SARIMAX 模型在差分移动自回归模型(ARIMA)的基础上加上季节(S, Seasonal)和外部因素(X, eX-

ogenous), 即在 ARIMA 模型的基础上加上周期性和季节性。ARIMA( $p, d, q$ )中, AR 是“自回归”,  $p$  是自回归项数; MA 是“滑动平均”,  $q$  是滑动平均项数,  $d$  是使之成为平稳序列所做的差分次数(阶数),  $L$  是滞后算子[10]。

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1-L)^d X_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (2)$$

SARIMAX( $p, d, q$ ) ( $P, D, Q, S$ )中,  $p$  表示趋势的自回归阶数;  $d$  表示趋势差分阶数;  $q$  表示趋势的移动平均阶数;  $P$  表示季节性自回归阶数;  $D$  表示季节性差分阶数;  $Q$  表示季节性移动平均阶数;  $S$  表示单个季节周期性的时间步长数。

由于时间序列分析方法要求所采用的观测数据之间有某种联系, 即未来的数据将会受过去数据发生的情况以及当时所处环境的影响, 所以对于时间序列数据而言过去的可以对未来的数据进行良好的预测。同时, 时间序列预测分析的方法要求数据应该严格遵循时间顺序, 使用者可依照各自的分析背景对时间间隔的数据进行抓取。

本文所用数据是: 由同一辆客车在同一路段在均匀时间间隔下记录的不同速度值所组成的观测数据, 完全符合 SARIMAX 模型初始要求, 故作者希望通过对车辆速度的定量分析进行建模, 从而对速度进行预测等一系列操作。

## 4.2. 车速预测模型

选取一组等时间间隔的数组, 以 140 号车在 1243 路段速度的三次样条插值数据为例, 建立 SARIMAX 模型。通过 ADF 检验的  $P$  值为  $0.01062 < 0.05$ , 故拒绝原假设, 可以认为序列是平稳的。接下来使用 BOX-Ljung 统计量判断白噪声, 得到  $P$  值为  $6.708e-36 < 0.05$ , 因此可以拒绝原假设, 即该序列通过了白噪声检验, 是白噪声序列。

最后进行时间序列趋势分解, 结果如图 3 所示, 由图 3 可知 Trend 数据具有明显的趋势性, Seasonal 数据具有明显的周期性, Resid 是剩余的部分, 可以认为是去除了趋势和季节性数据之后, 稳定的数据, 是模型所需要的[11]。

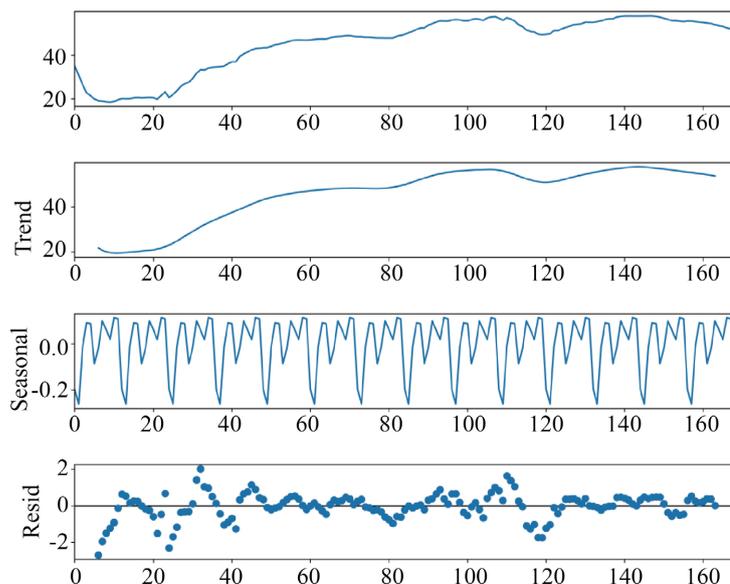


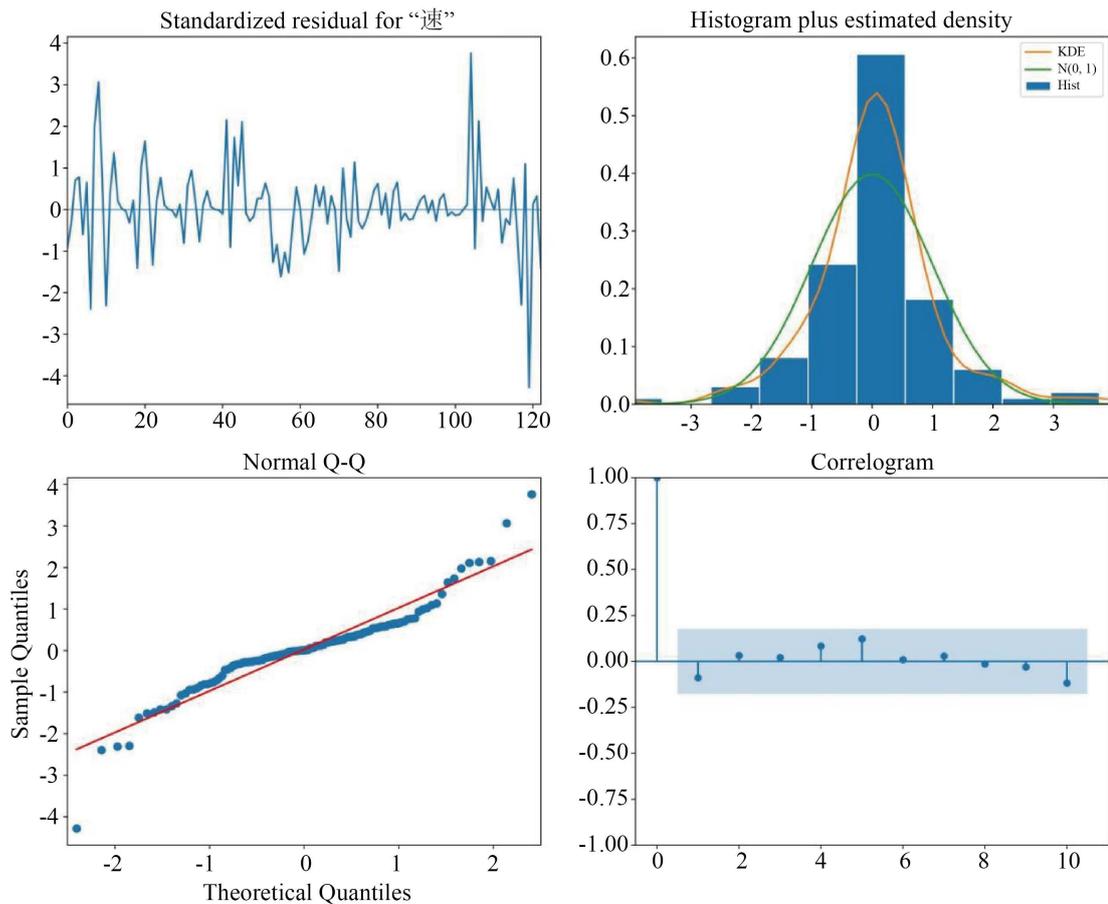
Figure 3. Time series trend breakdown of speed  
图 3. 速度的时间序列趋势分解图

按照差分次数确定阶数  $d$  和  $D$ , 按照 AIC (赤池信息准则) 和 BIC (贝叶斯信息准则) 确定最优模型, 得到最小 AIC 值为 597.949、最小 BIC 值为 609.198。即最优组合为:  $p = 1$ 、 $d = 1$ 、 $q = 1$ 、 $P = 0$ 、 $D = 1$ 、 $Q = 1$ 、 $S = 12$ 。随即进行参数估计, 结果如表 10 所示, 可知 ar.L1、ma.L1、ma.S.L12 和 sigma2 的  $P$  值都小于 0.05, 拒绝原假设, 即通过了显著性检验[11]。

**Table 10.** Parameter estimation results  
**表 10.** 参数估计结果

	coef	std err	$P >  z $	[0.025	0.975]
ar.L1	0.8359	0.181	0.000	0.481	1.191
ma.L1	-0.7154	0.198	0.000	-1.103	-0.327
ma.S.L12	-0.5825	0.078	0.000	-0.734	-0.430
sigma2	7.0057	0.523	0.000	5.980	8.031
SARIMAX Results			AIC	BIC	HQIC
Model: SARIMAX(1, 1, 1) × (0, 1, 1, 12)			597.949	609.198	602.518

模型诊断结果如图 4 所示:



**Figure 4.** Model diagnosis results  
**图 4.** 模型诊断结果

左上角为残差图，残差没有随时间的变化而变化，故通过了白噪声检验[12]；左下角为“q-q”图，残差的分布符合  $N(0, 1)$  的标准正态分布采样的线性趋势，这表示观测值具有良好的正态性[12]；右上角的图中红色 KDE 线与标准正态分布曲线( $N(0, 1)$ )行，这表明残差正常分布；右下角为残差自相关性图，表明时间序列残差与其本身的滞后版本具有低相关性。

对建立的模型进行验证预测，有助于了解 SARIMAX 模型预测的准确性。如图 5 所示，模型拟合的效果良好。

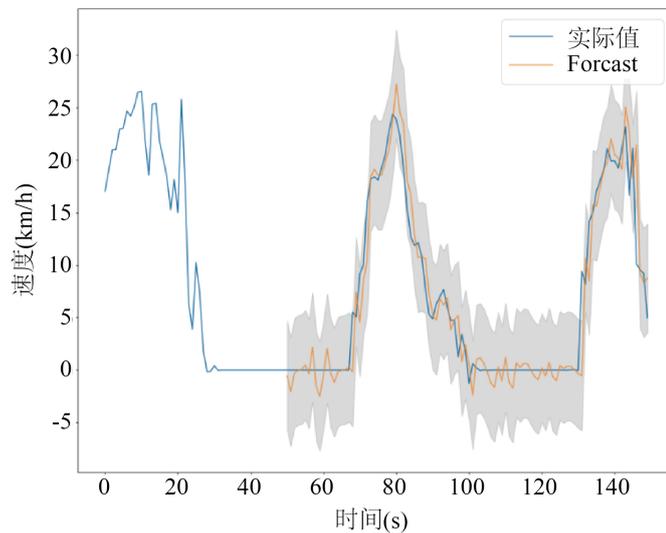


Figure 5. SARIMAX validation prediction results

图 5. SARIMAX 验证预测结果

通过观察诊断结果可知，模型较为理想，可以进行下一步预测[13]。最后利用最优模型(SARIMAX(1, 1, 1) (0, 1, 1, 12)模型)对未来汽车速度进行预测，得到的预测结果如图 6 示。

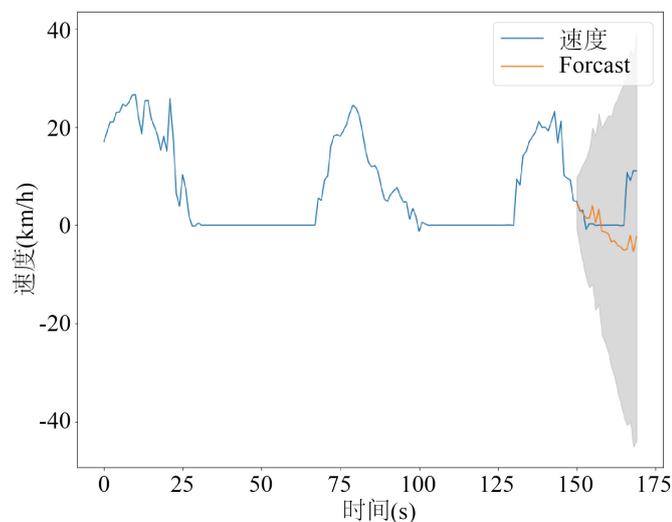


Figure 6. Forecast of future speed of 1243 section (low speed)

图 6. 1243 路段(低速)未来车速预测

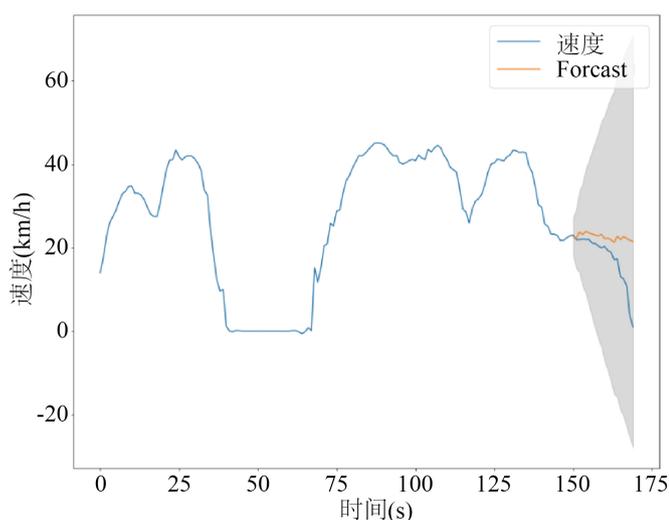
将预测结果与实际结果进行对比, 如表 11 所示, 对 15 个数据的预测中 SARIMAX 模型的准确率为 100%; 20 个速度的预测准确率为 80%。可见在低速工况下, SARIMAX 模型在中短期内的预测较为准确。

**Table 11.** Comparison between model prediction and actual results of Section 1243 (low speed)

**表 11.** 1243 路段(低速)模型预测与实际结果对比

编号	101	102	103	104	105	106	107	108	109	110
实际状态	A	A	A	A	A	A	A	A	A	A
预测状态	A	A	A	A	A	A	A	A	A	A
编号	111	112	113	114	115	116	117	118	119	120
实际状态	A	A	A	A	A	A	B	B	B	B
预测状态	A	A	A	A	A	A	A	A	A	A

同样利用 SARIMAX 模型对其他三个工况进行速度预测, 图 7 为通过 SARIMAX 模型得到的 155 号车在 1499 路段的速度预测。



**Figure 7.** Forecast of future speed of 1499 section (medium speed)

**图 7.** 1499 路段(中速)未来车速预测

将预测结果与实际结果进行对比, 如表 12 所示, 当利用 SARIMAX 模型预测 10 个速度数据时, 模型的准确率为 100%; 预测 15 个速度数据时, 准确率为 86.67%; 预测 20 个速度数据时, 准确率为 65%。因此在中速工况下, SARIMAX 模型在短期内的预测较为准确。

**Table 12.** Comparison between model prediction and actual results of 1499 section (medium speed)

**表 12.** 1499 路段(中速)模型预测与实际结果对比

编号	101	102	103	104	105	106	107	108	109	110
实际状态	C	C	C	C	C	C	C	C	C	C
预测状态	C	C	C	C	C	C	C	C	C	C

Continued

编号	111	112	113	114	115	116	117	118	119	120
实际状态	C	C	C	B	B	B	B	B	A	A
预测状态	C	C	C	C	C	C	C	C	C	C

图 8 为通过 SARIMAX 模型得到的 351 号车在 808 路段的速度预测。

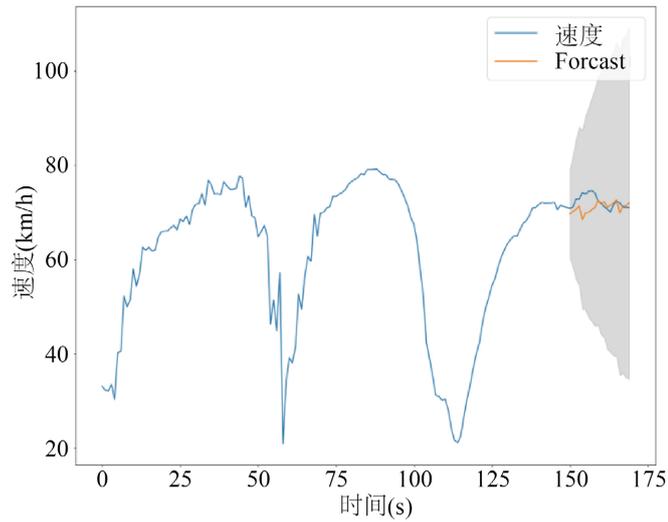


Figure 8. Forecast of future speed of 808 section (medium and high speed)

图 8. 808 路段(中高速)未来车速预测

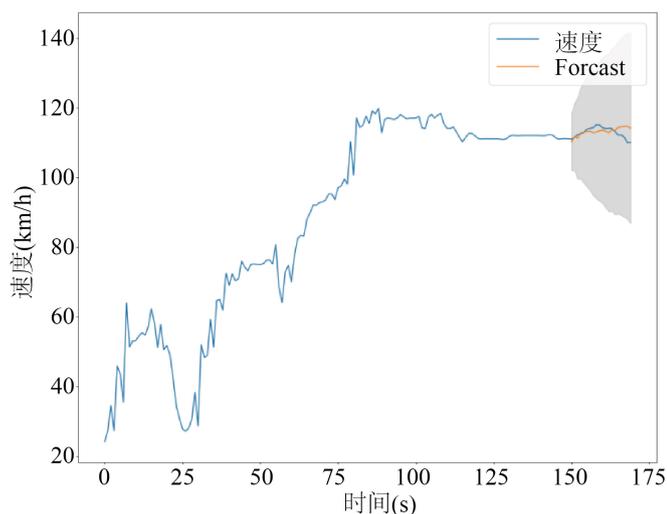
将预测结果与实际结果进行对比, 如表 13 所示, 当利用 SARIMAX 模型预测 5 个速度数据时, 模型预测的准确率为 40%; 预测 10 个速度数据时, 准确率为 30%; 预测 15 个速度数据时, 准确率为 53.33%; 预测 20 个速度数据时, 准确率为 65%。在中高速工况下, SARIMAX 模型在长期内的预测较为准确, 在中短期的速度预测与实际值有所偏差。

Table 13. Comparison between model prediction and actual results of 808 section (medium and high speed)

表 13. 808 路段(中高速)模型预测与实际结果对比

编号	101	102	103	104	105	106	107	108	109	110
实际状态	H	H	I	I	I	I	I	I	I	H
预测状态	H	H	H	H	H	H	H	H	H	H
编号	111	112	113	114	115	116	117	118	119	120
实际状态	H	H	H	H	H	H	H	H	H	H
预测状态	H	H	H	H	H	H	H	H	H	H

图 9 为通过 SARIMAX 模型得到的 147 号车在 712 路段的速度预测。



**Figure 9.** Speed prediction of 712 section (high speed)  
**图 9.** 712 路段(高速)车速预测

将预测结果与实际结果进行对比,如表 14 所示,当利用 SARIMAX 模型预测 5 个速度数据时,模型预测的准确率为 100%;预测 10 个速度数据时,准确率为 60%;预测 15 个速度数据时,准确率为 60%;预测 20 个速度数据时,准确率为 45%。可见在高速工况下, SARIMAX 模型在短期内预测较为准确。

**Table 14.** Comparison between predicted results and actual results

**表 14.** 预测结果与实际结果对比

编号	101	102	103	104	105	106	107	108	109	110
实际状态	N	N	N	N	N	O	O	O	O	O
预测状态	A	N	N	N	N	N	N	N	N	N
编号	111	112	113	114	115	116	117	118	119	120
实际状态	O	O	O	O	N	N	N	N	N	N
预测状态	D	O	O	O	N	O	O	O	O	O

## 5. 对比分析

不同工况下,利用马尔可夫模型和 SARIMAX 模型分别预测未来 5 个、10 个、15 个和 20 个车辆速度,并与实际值对比,得到的准确率如表 15。

**Table 15.** Comparison of Markov Model and SARIMAX Model

**表 15.** 马尔可夫模型和 SARIMAX 模型的比较

	低速		中速		中高速		高速	
	马尔可夫	SARIMAX	马尔可夫	SARIMAX	马尔可夫	SARIMAX	马尔可夫	SARIMAX
5 个	100%	100%	80%	100%	40%	40%	100%	100%
10 个	100%	100%	40%	100%	30%	30%	50%	60%
15 个	100%	100%	26.67%	86.67%	53.33%	53.33%	40%	60%
20 个	80%	80%	20%	65%	40%	65%	55%	45%

在低速工况下, 利用马尔可夫模型和 SARIMAX 模型对未来车速进行预测的效果相差无几, 且在中短期内的预测效果都较好; 在中速工况下, SARIMAX 模型的预测效果优于马尔可夫模型, 且 SARIMAX 模型在中短期内预测的准确率较高, 而马尔可夫模型预测的效果不尽如人意; 在中高速工况下, 马尔可夫模型和 SARIMAX 模型预测的准确率都不高, 相比之下, SARIMAX 模型的长期预测效果优于马尔可夫模型; 在高速工况下, SARIMAX 模型和马尔可夫模型在短期内的预测效果都较好, 但是中长期的预测准确率较低, 相比较而言马尔可夫模型在长期的预测效果稍优于 SARIMAX 模型。

通过对比分析发现, SARIMAX 模型在低速、中速和高速工况下的短期预测效果较优, 在低速工况下的长期预测准确率也较高, 在中速和中高速工况下的长期预测效果优于马尔可夫模型; 马尔可夫模型在低速和高速工况下的短期预测效果较优, 在低速工况下的长期预测效果较好, 在高速工况下的长期预测效果稍优于 SARIMAX 模型。

在拟合车辆速度方面, 马尔可夫模型侧重于以区间进行速度分类, 通过大数统计来对下一时间段的车辆速度进行拟合。马尔可夫模型是一种对历史数据进行统计分析, 发现统计规律从而进行预测的模型, 它对于临时改变状况的因素考虑不足, 比较适合预测路况较为平稳, 意外因素较少的车况。

SARIMAX 模型侧重于研究车辆速度的历史数据并进行拟合, 每一个时刻的速度只与上一时刻有关。SARIMAX 模型基于平稳的时间序列数据, 将时间序列数据代入模型进行预测, 一个时刻估计值的差分与上一个时刻的预测误差有关, 即对预测的车速的关联度较高, SARIMAX 模型依赖于相邻上一时刻的汽车行驶状态, 更符合车辆实际行驶状态的工况。

## 6. 结论

本文基于马尔可夫模型和 SARIMAX 模型, 提出将二者引入汽车速度预测模型。通过实际操作研究发现, 在低速工况下, 对车速进行短期预测抑或是长期预测, SARIMAX 模型和马尔可夫模型都是不错的选择; 在高速工况下, 对未来车速进行长期预测应选择马尔可夫模型, 短期预测两种模型都能获得较好的效果; 对于中速和中高速工况, 选择 SARIMAX 模型对车速进行长期预测能获得更好的效果; 在中速工况下对车速进行短期预测, 选择 SARIMAX 模型较为合适; 在中高速工况下, 利用 SARIMAX 模型和马尔可夫模型进行短期预测的效果都不太理想, 可考虑选用其他模型对车辆速度进行预测。本文的实证研究表明不同工况、不同预测长度及模型的选取对车辆速度预测的精度会产生很大的影响。

## 基金项目

2021 年度国家级大学生创新创业训练计划(202111058035)。

## 参考文献

- [1] 李文茹. 城市汽车运行速度分布预测模型研究[J]. 吉林大学, 2018(6): 2-5.
- [2] 管珊珊, 张益农. 汽车速度预测技术研究发展[J]. 计算机科学, 2017(10): 1-3.
- [3] 解少博, 刘通, 李会灵. 基于马尔科夫链的并联 PHEB 预测型能量管理策略研究[J]. 汽车工程, 2018, 40(8): 871-877+911.
- [4] 袁焕涛, 赵红, 潘广纯. 基于马尔科夫的混合动力汽车行驶车速预测研究[J]. 智能城市, 2021, 7(3): 18-22.
- [5] 张良力, 祝贺, 吴超仲. 基于 ARMA 预测模型的交叉口车辆碰撞风险评估[J]. 交通运输系统工程与信息, 2015, 15(5): 239-245.
- [6] Guo, J., He, H. and Sun, C. (2019) ARIMA-Based Road Gradient and Vehicle Velocity Prediction for Hybrid Electric Vehicle Energy Management. *IEEE Transactions on Vehicular Technology*, **68**, 5309-5320. <https://doi.org/10.1109/TVT.2019.2912893>
- [7] Geunseob, O.H., LeBlanc, D.J. and Peng, H. (2020) Vehicle Energy Dataset (VED): A Large-Scale Dataset for Vehicle

---

Energy Consumption Research. *IEEE Transactions on Intelligent Transportation Systems*, **23**, 3302-3312.  
<https://doi.org/10.1109/TITS.2020.3035596>

- [8] 陈华友, 周礼刚, 刘金培. 统计预测与决策[M]. 北京: 科学出版社, 2018(6): 46-60.
- [9] 夏丽. 基于 ARIMA 模型及回归分析的区域用电量预测方法研究[D]: [硕士学位论文]. 南京: 南京理工大学, 2013: 2-6.
- [10] 孙义, 周陇陇. 基于 Python 的金融时间序列 ARIMA 模型教学[J]. 现代信息科技, 2021, 5(10): 192-195.
- [11] 涂锦, 冷正兴, 刘丁毅. 基于 EMD 和神经网络的非线性时间序列预测方法[J]. 统计与决策, 2020(8): 1.
- [12] 李潇俊, 唐攀. 基于技术分析、基本面分析和深度学习的股价预测[J]. 统计与决策, 2022(2): 2-5.
- [13] 姜向荣. 时间序列分析与预测[M]. 北京: 科学出版社, 2020: 20-29.