

铁路客流量规律分析报告

佛山科学技术学院

钱培淳、谢雷、黄晟

一、铁路客流规律研究背景及意义

长期以来，凭借着低廉的价格，四通八达的线路和极高的安全性，中国普通铁路成为广大老百姓长途出行的首选。

铁路部门为了保持市场的竞争力，实现利润的最大化，需要了解日常铁路客运流量、淡旺季变动指数、冷热门线路的具体情况，而其中对客流的充分了解和分析是准确把握市场的首要条件，因此铁路客流分析的研究也成为铁路客运服务需要重点研究的方向。

该作品通过某铁路公司客运专线客流量数据为优化客运专线运力资源配置提供研究依据。

1. 数据处理与分析

利用某铁路公司客运专线采集的客流量数据信息，采用数据挖掘技术，通过数据筛选、清理、构建、整理合并、规范化等步骤，从客流结构、客流出行时间、铁路对客流的吸引程度等方面对该铁路客流规律进行分析，以及通过车站所属地区的气象信息分析天气状况、气温、风力风向与铁路客流量的相关关系。

2. 数据预测

总结客流量规律之后，根据分析气象因素与铁路客流量的相关性，建立对应平稳时间序列 ARMA 模型和非平稳序列 ARIMA 模型对客流量进行预测，分别利用 2015 年客流量数据对 2016 年的 1-2 月客流量进行预测验证，并对 2016 年 3 月之后的客流量做了预测，结果表明预测数据和实际相近，预测结果较为精准。

3. 程序设计

- 1) 设计数据库软件 SQL Server 2008 实现数据分析。
- 2) 使用 CSharp 实现数据表中数据的自动导入功能。
- 3) 通过矩阵实验室 (matlab) 开发平台可视化展现不同时间段不同车次的铁路客流量变化。
- 4) 根据不同站点 15 个月的客流规律，使用 R 语言针对不同站点建立相对应的平稳、非平稳时间序列模型进行站点客流量分析预测。

二、步骤流程与数据处理框架

1. 总体流程

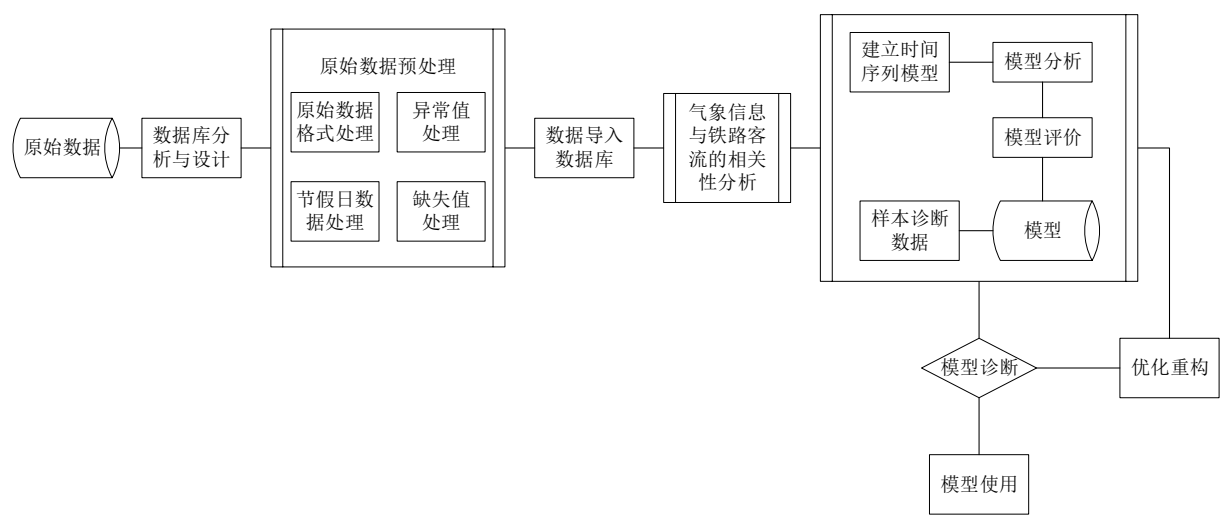


图 1 总体流程图

本文对铁路客流规律进行了深入分析和研究，建立了平稳时间序列 ARMA 模型、非平稳序列 ARIMA 模型，制作了基于各站点不同时间段客流量变化的铁路客流量系统。

步骤一：针对问题进行相关性分析，确定建模思路，从原始数据集中选取样本数据。

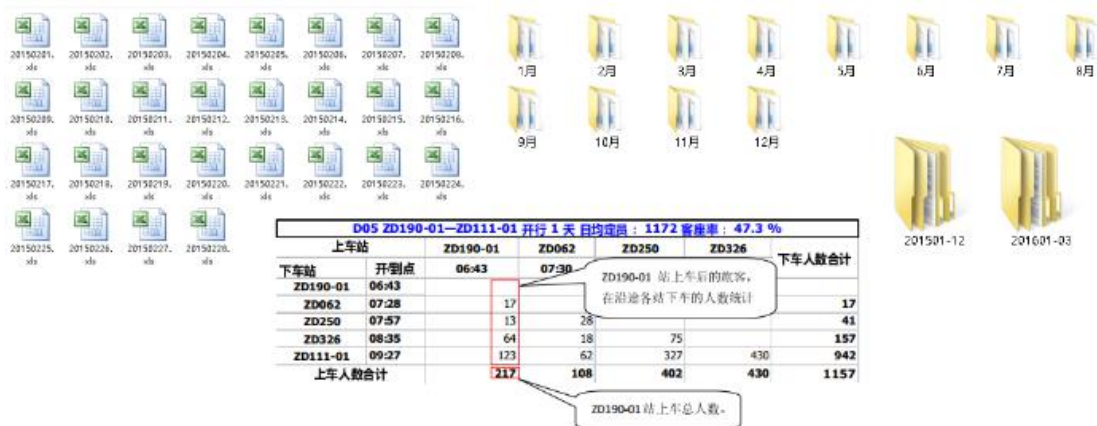
步骤二：数据预处理。

步骤三：气象信息与客流量的相关性分析。

步骤四：建立时间序列数学模型。

步骤五：基于各站点不同时间段客流量变化的铁路客流量可视化展现。

2. 原始数据文件结构



450 天每个车次始发日期、日均定员、所经站点、客座率，以及在不同时间点的站点处上下车人数以及总客流量（备注：文件有 excel 和 html 表格形式）

数据来源于铁路公司记录（为确保铁路公司隐私问题，各站点名称均由字母及编号表示）。保障了本次数据分析的真实性、可靠性以及完整性。大幅度减少误差程度以及对数据分析的片面性。

3. 数据库

3.1 数据库分析与设计

采用数据库的方式对采集数据进行存储管理，可便于提取重要信息、分析数据，并且可用于铁路部门对车站客流数据的管理，包括各类数据信息的添加、编辑、查询和删除等。

使用 SQL Server 2008 开发数据库，可直接对文件进行集中管理，开发时还可以根据需要向其中添加或新建数据库、表、表单和报表等。在各个表中输入数据记录，并在数据库设计器中根据表的共同字段“地区”、“日期”、“站点”、“车

次”建立表之间一对一的关系。

3.2 功能模块设计

- ① 地区管理模块：地区名。
- ② 站点信息模块：地区、站点。
- ③ 天气状况记录模块：地区、日期、起始天气、转换后天气、最低气温、最高气温、起始风力风向、转换后风力风向。
- ④ 车次记录模块：车次、日期、出发站、终点站、日均定员、实际乘车人数、客座率。
- ⑤ 站点记录模块：车次、日期、当前站点、车次抵达时间、发车时间、上车人数、下车人数。
- ⑥ 客流信息模块：车次、日期、上车站、下车站、乘客数。

3.3 数据库数据查询通道

地址：win.dimago.cn:1433

账号：RailWayTest

密码：railway

（可使用 SSMS 查看数据库）

3.4 数据关系图

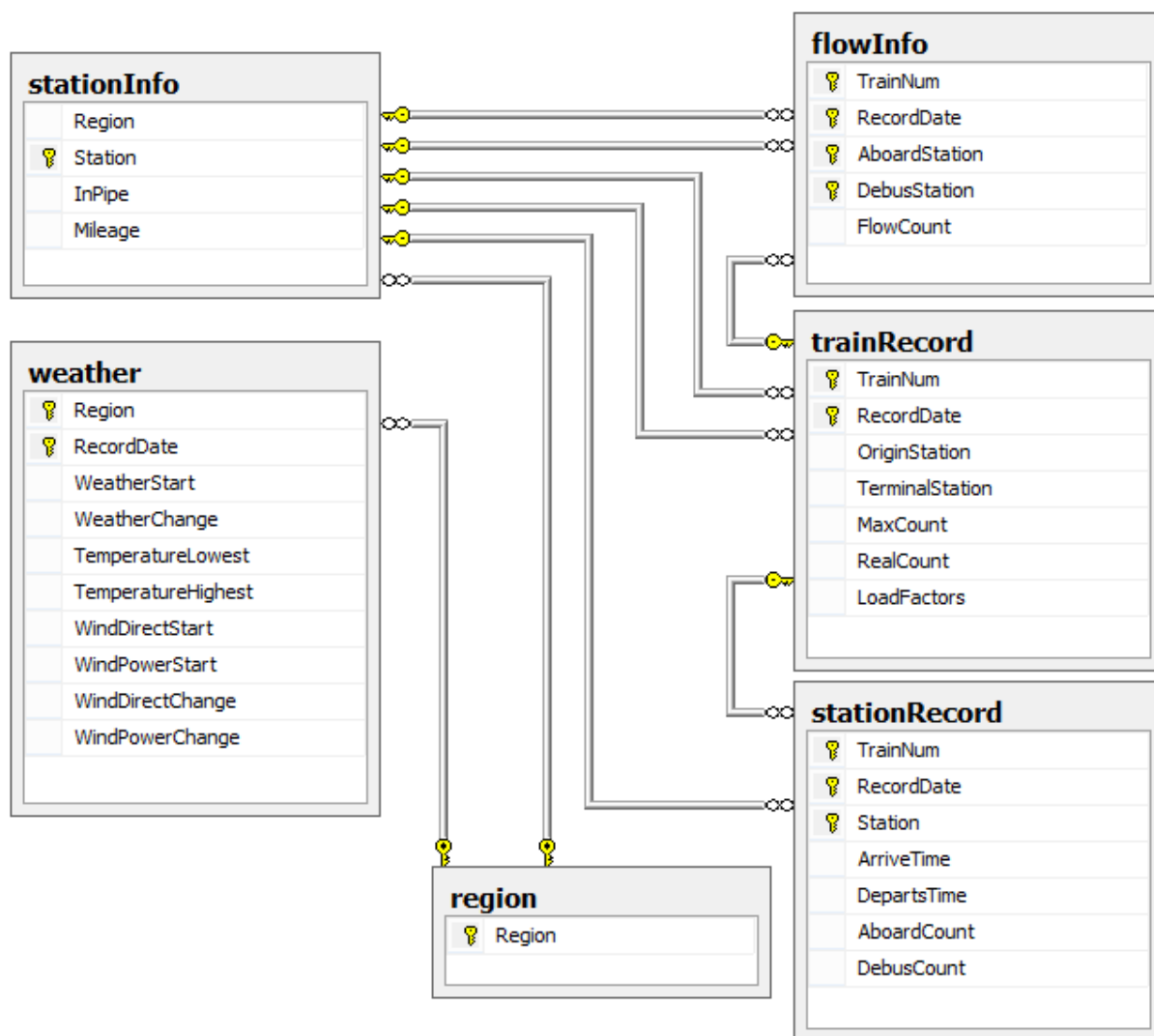


图 2 数据关系图

4. 数据提取与导入

4.1 数据提取器

数据提取采用 C#编写提取器进行提取。数据提取器效果如图 3 所示。



图 3 数据导入器

4.2 数据提取工作流程

首先进行数据的预处理，对于原始的旅客列车梯形密度表，将列车车次、开车日期、出发站点、出发时间、终点站、终到时间、客座率提取至 trainRecord.csv 文件中，将途径车站、到达时间、出发时间对应车次信息提取至 stationRecord.csv 文件中，最后将站点间的客流信息提取出来对应车次信息和起始站点置入 flowInfo.csv 中。

4.3 数据导入工作流程

4.3.1 将提取的数据导入数据库

将 trainRecord.csv、stationRecord.csv、flowInfo.csv 中的数据上传到服务器中对应的 trainRecord、stationRecord、flowInfo 表中，由数据库的触发器通过 flowInfo 的客流人数计算出每个列车的载客人数和在各站

点的上下车人数。

4.3.2 将天气情况导入数据库

读取天气情况记录文件 Weather.csv，并将内容导入数据库中的 weather 表中。

5. 数据预处理

数据预处理主要包括缺失值处理、异常值处理、节假日数据处理、数据提取文件格式修正等。

5.1 缺失值处理

由附件数据集中的历史数据可知，特别是文件数据抽取过程中，发现存在缺失的现象，根据缺失值处理原则，为确保建模数据的有效性，可直接去除这些值，在进行相关计算时，不列入其中，不会影响分析建模。特殊情况：针对发车时间迟于到站时间的情况，是由于缺失车次时间的缘故，填补空值以确保数据的完整性。

5.2 异常值处理

在原始样本数据中，存在部分异常值的情况，如列车进站后只有发车时间没有到站时间，以及一些超出指标阈值范围的数据，对于这类异常数据，当成缺失值处理。

5.3 节假日数据处理

节假日的铁路客流量同工作日比起来，会明显偏高，为了确保不同日期类型的计量数据在一定时间周期内具有可比性和连贯性，需要对这些特殊类型日期的数据进行修正处

理，处理规则是：区分开非节假日和节假日，分别建立对应的时间序列模型对未来客流量进行预测，例如清明几天的铁路客流量是通过抽取已知数据中的节假日客流信息之后建立相应的 ARMA 模型进行预测。

5.4 数据提取文件格式修正

将所有 xls 和 3 月份的 html 数据文件通过执行程序统一转换为 csv 文件处理。

6. 铁路客流量可视化展现

该系统 exe 可在安装有 matlab 的机器上运行，若在没有安装 matlab 的计算机上，可先安装 matlab 运行环境 MCRInstaller，安装完成即可正常使用。

软件的界面设计，左方为演示区、右方为操作区。主要功能包括：中国铁路路线图查看、日期选择、车次选择、载入客流信息库、乘客数表示方式选择（颜色或线条）、演示程序、查看测试样例、重置系统、退出系统等。

本系统提供中国铁路路线图，可使用菜单栏中的缩放，移动功能进行相关查看。日期选择和车次选择中有选择菜单，可针对不同日期不同车次进行站点客流演示。本系统的设计针对用户不同需求提供了乘客数表示方式：颜色或线条，根据乘客数的大小，颜色深浅或线条粗细会发生相应变化，用户可以更为直观地查看。

本系统主要功能是在一定时间内对铁路公司管内各站

点乘客数变化进行动态演示。说明（红色箭头为列车）：列车从站点出发时会有乘客上车，抵达站点时有乘客上车也有乘客下车。在经过一个站点时，其乘客数就可能会发生改变。

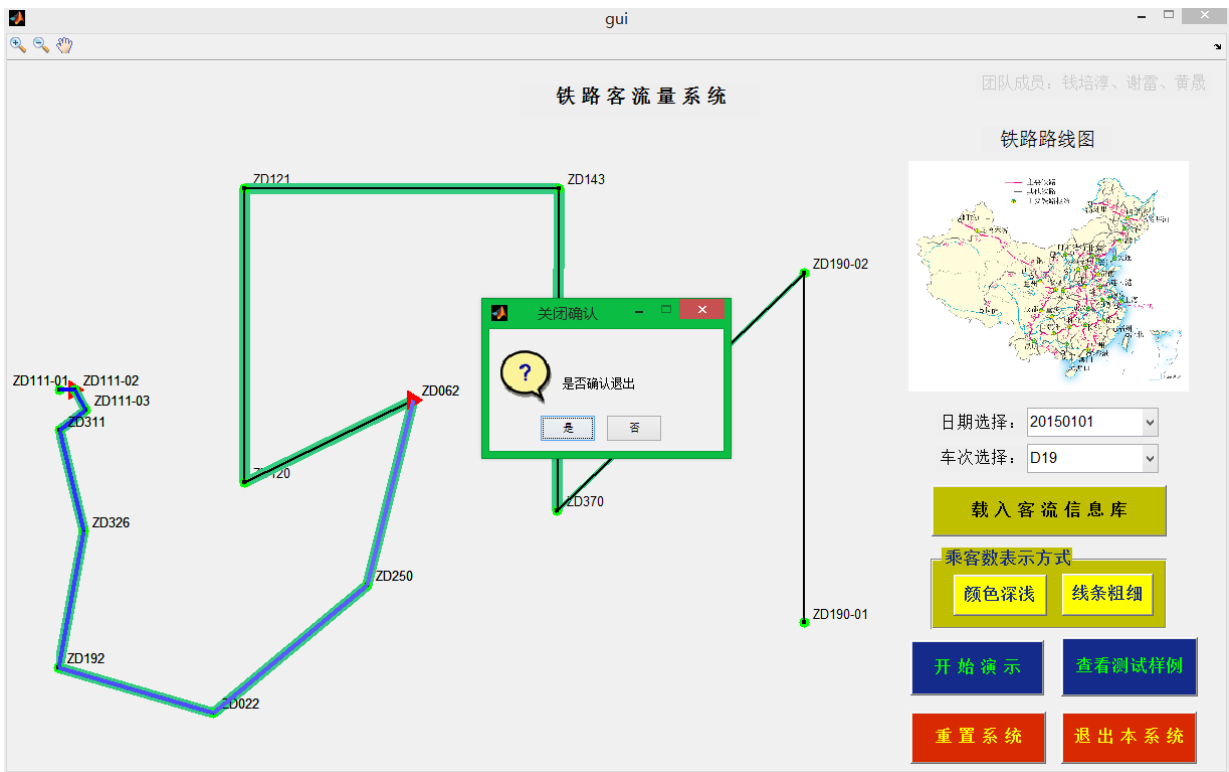


图 4 铁路客流量系统

三、数据分析

1. 气象信息与客流量的相关性分析

天气因素可能影响铁路的客流量。这里采用寒风指数 WCI 与客流量的相关性进行分析，验证客流量是否收到天气因素的影响。

1.1 WCI 公式

风寒指数（WCI）是表征寒冷环境条件下，风速与气温对裸露人体的影响，是人类对低温和风的一种感觉程度。物理意义是指皮肤温度为 33℃时，体表单位面积的散热量。其

计算公式如下。公式中，t 表示摄氏气温(℃)，V 表示风速(m/s)。

$$WCI = (33 - t)(9.0 + 10.9\sqrt{V} - V)$$

WCI 与人的感觉状况对应关系如表 1 所示。

范围	人的感觉状况	级别
> 1000	很冷风	e
800 - 1000	冷风	d
600 - 800	稍冷风	c
300 - 600	凉风	b
200 - 300	舒适风	A
50 - 200	暖风	B
-80 - 50	皮肤不明想风	C
-160 - -80	皮肤赶热风	D
< -160	皮肤不适风	E

表 1 WCI 与人的感觉对应图

1.2 天气参数值计算公式

在本次分析中，要计算天气情况与客流量的相关性系数，则需要通过 WCI 计算出天气参数值。天气参数值与人体舒适程度呈负相关，天气参数值越大，相对人体舒适程度差。

$$\text{天气参数值} = |WCI - 300|$$

1.3 分析过程

1) 首先，从数据库中提取出管内车站所在城市每天的客流量数据。

城市客流量 = \sum 当天该城市所有站点上车人数 + 当天该城市所有站点下车人数

2) 然后提取出这些城市每一天的天气参数，通过天气参数计算出 WCI 指数。

温度参数取最高温度与最低温度之和的平均值。风速取前部分风速的最小值（风速 ≤ 3 时取 2）与后部分的最大值之和的平均值。

3) 对 WCI 指数进行预处理，得到天气参数值。

4) 将 WCI 处理值与客流量做相关性系数计算，计算结果如下。

	ZD111天气参数值	ZD111客流			ZD250天气参数值	ZD250客流
ZD111天气参数值	1			ZD250天气参数值	1	
ZD111客流	-0.290392084	1		ZD250客流	-0.247661324	1
	ZD190天气参数值	ZD190客流			ZD326天气参数值	ZD326客流
ZD190天气参数值	1			ZD326天气参数值	1	
ZD190客流	-0.535701523	1		ZD326客流	-0.241288898	1

表 2 管内车站天气参数值与客流量相关性系数计算结果

2. 站点客流信息统计

2.1 各车站客流量统计

2.1.1 统计过程

1) 从数据库中分别导出各车站 2015-01-01 至 2016-03-20 期间每日的客流量（上车人数与下车人数之和）。

2) 以日期为横坐标、客流量为纵坐标，在同一张图上绘制各车站的客流量折线统计图。

3) 利用 excel 的统计插件对站点间两两做客流量的相关

性系数计算，得出相关性系数表。

2.1.2 统计结果

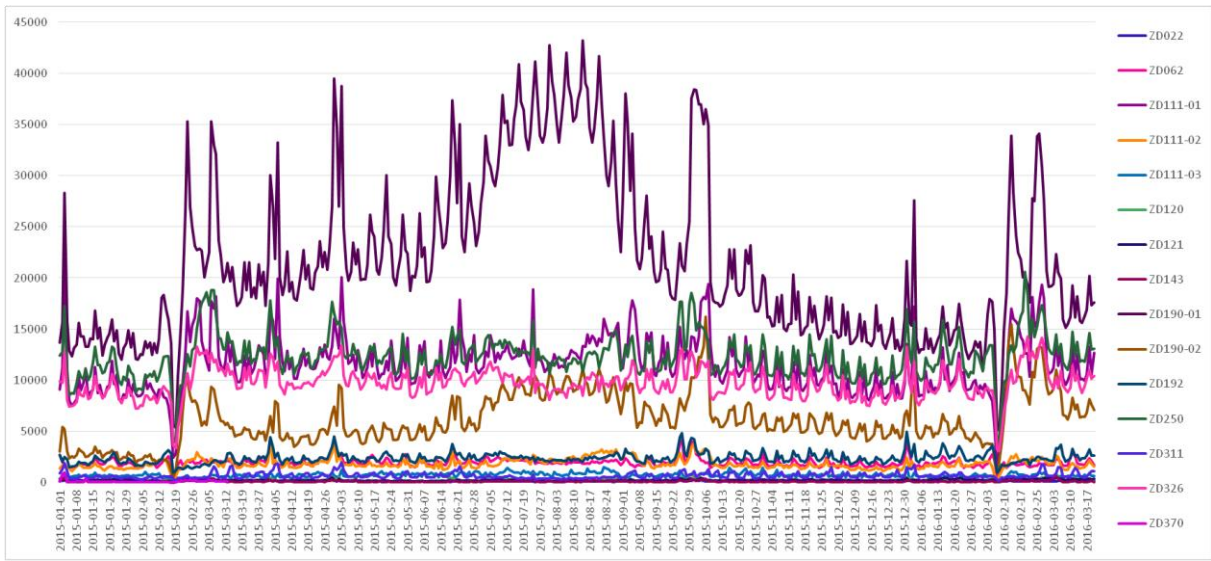


图 5 客流量折线统计图

图 5 是 2015-01-01 至 2016-03-20 期间各车站客流量折线统计图。本图以日期为横坐标，客流量为纵坐标，以研究各车站客流量变化情况，主要用于分析各车站的客流量变化规律。

	ZD022	ZD062	ZD111-01	ZD111-02	ZD111-03	ZD120	ZD121	ZD143	ZD190-01	ZD190-02	ZD192	ZD250	ZD311	ZD326	ZD370	总客流
ZD022	1															
ZD062	0.472303	1														
ZD111-01	0.351596	0.738481	1													
ZD111-02	0.116033	0.771182	0.655772	1												
ZD111-03	0.060091	0.28769	0.633225	0.215463	1											
ZD120	0.269668	0.665113	0.554326	0.655175	0.244711	1										
ZD121	0.336353	0.294572	0.341477	0.239219	0.319976	0.627435	1									
ZD143	0.284482	0.601116	0.53582	0.580306	0.222649	0.807181	0.410456	1								
ZD190-01	0.079284	0.753366	0.708237	0.756161	0.285993	0.435881	-0.00121	0.391424	1							
ZD190-02	0.073614	0.691063	0.681622	0.696208	0.369117	0.671838	0.469365	0.513342	0.770668	1						
ZD192	0.442939	0.792535	0.730776	0.627082	0.504519	0.59658	0.371004	0.547269	0.542464	0.607591	1					
ZD250	0.437013	0.805823	0.946943	0.709193	0.571674	0.656202	0.436828	0.586202	0.676054	0.697468	0.78366	1				
ZD311	0.613909	0.625923	0.625641	0.376043	0.405073	0.487845	0.38095	0.492623	0.244547	0.296718	0.788901	0.676195	1			
ZD326	0.428367	0.781893	0.959185	0.675328	0.571268	0.595374	0.375006	0.553182	0.694667	0.674489	0.757008	0.9743	0.672511	1		
ZD370	0.237154	0.364958	0.387593	0.161195	0.191945	0.879408	0.396668	0.882661	0.230415	0.199881	0.416739	0.342154	0.575127	0.386495	1	
总客流	0.290817	0.860316	0.926689	0.803849	0.503903	0.633599	0.297836	0.570128	0.893502	0.834741	0.763452	0.923575	0.545252	0.925526	0.360622	1

表 3 客流量相关性系数表

表 3 是 2015-01-01 至 2016-03-20 期间各车站日客流量相关性系数表。表中数据为对应两个站点在 2015-01-01 至

2016-03-20 期间日客流量变化的相关性系数。

2.2 各车站上车人数统计

2.2.1 统计过程

1) 从数据库中分别导出各车站 2015-01-01 至 2016-03-20 期间每日的上车人数。

2) 以日期为横坐标、上车人数为纵坐标，在同一张图上绘制各车站的上车人数折线统计图。

3) 利用 excel 的统计插件对站点间两两做上车人数的相关性系数计算，得出相关性系数表。

2.2.2 统计结果

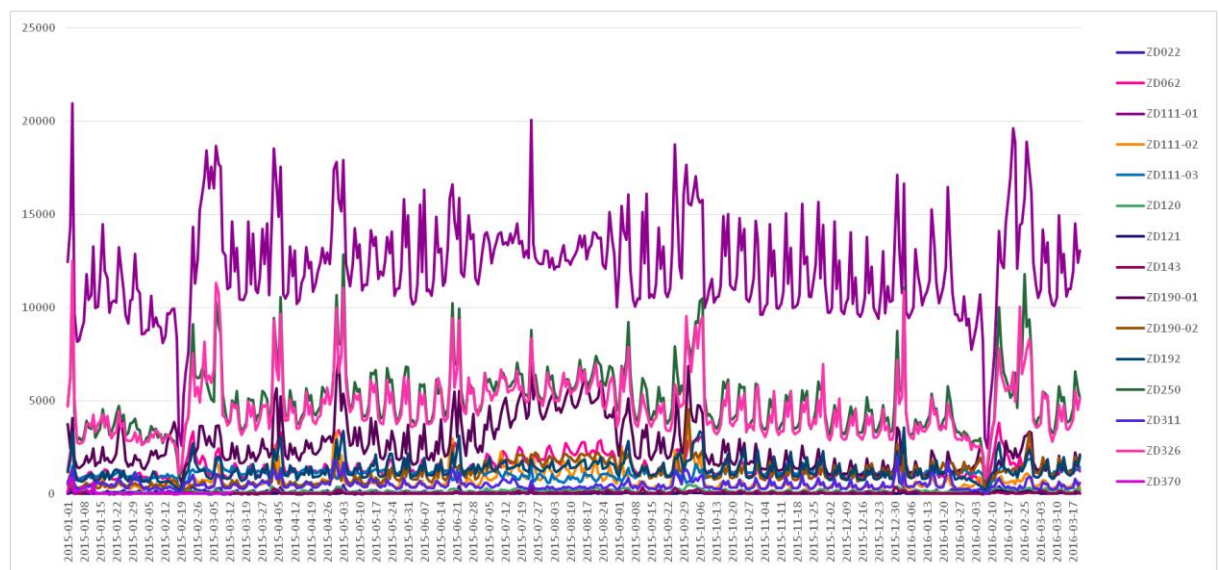


图 6 上车人数折线统计图

图 6 是 2015-01-01 至 2016-03-20 期间各车站上车人数折线统计图，该图以日期为横坐标，站点上车人数为纵坐标，以研究各车站上车人数变化情况。

2.2.3 相关性系数

	ZD022	ZD062	ZD111-01	ZD111-02	ZD111-03	ZD120	ZD121	ZD143	ZD190-01	ZD190-02	ZD192	ZD250	ZD311	ZD326	ZD370	总人数
ZD022	1															
ZD062	0.671419	1														
ZD111-01	0.429134	0.554961	1													
ZD111-02	0.108238	0.538603	0.56646	1												
ZD111-03	0.343945	0.067235	0.535744	-0.08789	1											
ZD120	0.305768	0.644657	0.521707	0.612731	0.073702	1										
ZD121	0.554107	0.548302	0.278595	0.198641	0.226418	0.506093	1									
ZD143	0.201197	0.424034	0.451385	0.43848	0.122829	0.726942	0.237714	1								
ZD190-01	0.299257	0.684027	0.595007	0.608384	-0.04515	0.458841	0.155143	0.338213	1							
ZD190-02	0.172565	0.527637	0.500423	0.578592	0.105205	0.612826	0.408826	0.371151	0.594265	1						
ZD192	0.709586	0.811205	0.681072	0.466298	0.351629	0.580189	0.477076	0.49957	0.55309	0.538243	1					
ZD250	0.639209	0.822616	0.827093	0.608956	0.381151	0.62166	0.47918	0.490894	0.681527	0.566032	0.870056	1				
ZD311	0.426445	0.428682	0.654092	0.52871	0.370156	0.427846	0.20392	0.440088	0.353183	0.308003	0.630283	0.660521	1			
ZD326	0.631287	0.790548	0.826838	0.62703	0.347463	0.594161	0.438134	0.445664	0.673769	0.528923	0.838202	0.953241	0.635149	1		
ZD370	0.20309	0.232662	0.22827	0.129491	0.01567	0.829032	0.436584	0.876488	0.123899	-0.00244	0.348654	0.27249	0.309059	0.284699	1	
总人数	0.566739	0.802389	0.906013	0.691756	0.366039	0.657264	0.419807	0.516598	0.769972	0.646644	0.847534	0.963023	0.675448	0.953515	0.291771	1

表 4 上车人数相关性系数表

表 4 是 2015-01-01 至 2016-03-20 期间各车站日上车人数相关性系数表。表中数据为对应两个站点在 2015-01-01 至 2016-03-20 期间日上车人数变化的相关性系数。

2.3 各车站下车人数统计

2.3.1 统计过程

1) 从数据库中分别导出各车站 2015-01-01 至 2016-03-20 期间每日的下车人数。

2) 以日期为横坐标、客流量为纵坐标，在同一张图上绘制各车站的下车人数折线统计图。

3) 利用 excel 的统计插件对站点间两两做下车人数的相关性系数计算，得出相关性系数表。

2.3.2 统计结果

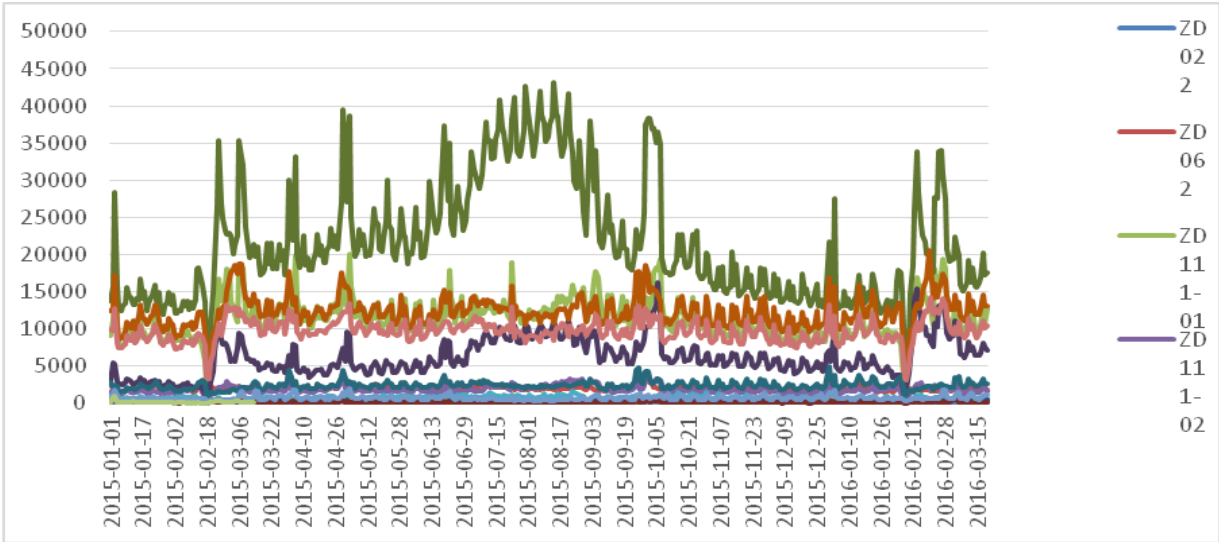


图 7 下车人数折线统计图

图 7 是 2015-01-01 至 2016-03-20 期间各车站下车人数折线统计图，该图以日期为横坐标，站点下车人数为纵坐标，以研究各车站下车人数变化规律。

2.3.3 相关性系数

	ZD022	ZD062	ZD111-01	ZD111-02	ZD111-03	ZD120	ZD121	ZD143	ZD190-01	ZD190-02	ZD192	ZD250	ZD311	ZD326	ZD370	总计
ZD022	1															
ZD062	0.398033	1														
ZD111-01	-0.05559	0.384596	1													
ZD111-02	-0.05397	0.520911	0.696287	1												
ZD111-03	-0.29298	0.278972	0.492514	0.410109	1											
ZD120	0.129949	0.5125	0.434846	0.499549	0.161644	1										
ZD121	0.182911	0.239078	0.239751	0.183812	0.053124	0.660679	1									
ZD143	0.15399	0.415283	0.460236	0.4906	0.095029	0.753114	0.432362	1								
ZD190-01	-0.21866	0.397786	0.705825	0.700897	0.465703	0.324506	-0.08491	0.376944	1							
ZD190-02	-0.26792	0.270576	0.698391	0.665541	0.386016	0.585091	0.37074	0.505988	0.778228	1						
ZD192	0.305086	0.780556	0.366272	0.455125	0.34932	0.472592	0.30873	0.326236	0.293652	0.289038	1					
ZD250	0.201199	0.653058	0.741829	0.621574	0.349991	0.626692	0.448589	0.532563	0.499993	0.557521	0.60628	1				
ZD311	0.363747	0.371605	0.47418	0.182382	0.060564	0.333696	0.274186	0.326573	0.111542	0.21798	0.519476	0.479321	1			
ZD326	0.149048	0.595214	0.800857	0.578438	0.382226	0.525124	0.382193	0.450526	0.506342	0.540466	0.554264	0.928429	0.502746	1		
ZD370	-0.0678	0.128897	0.520818	0.15123	0.145104	0.88609	0.398354	0.819363	0.404162	0.418999	-0.02068	0.372949	0.638332	0.380266	1	
总计	-0.09114	0.542413	0.879543	0.792817	0.510207	0.534612	0.195367	0.522901	0.916589	0.850644	0.481097	0.771339	0.343877	0.773372	0.465214	1

表 5 下车人数相关性系数表

表 5 是 2015-01-01 至 2016-03-20 期间各车站日下车人数相关性系数表。表中数据为对应两个站点在 2015-01-01 至 2016-03-20 期间日下车人数变化的相关性系数。

2.4 不同时间段客流量统计

2.4.1 统计过程

1) 从数据库中导出 0-24 小时，每个小时内全部车站在 2015-01-01 至 2016-03-20 期间每一天的客流量总和。

2) 以 0-23 时为横坐标，客流量之和为纵坐标，绘制在每日不同的时间段客流量变化规律的折线统计图。

3) 从数据库中分别导出各车站每个小时内在 2015-01-01 至 2016-03-20 期间每一天的客流量总和。

4) 以 0-23 时为横坐标，客流量之和为纵坐标，在同一张图中绘制个车站在每日不同的时间段客流量变化规律的面积图。

2.4.2 统计结果

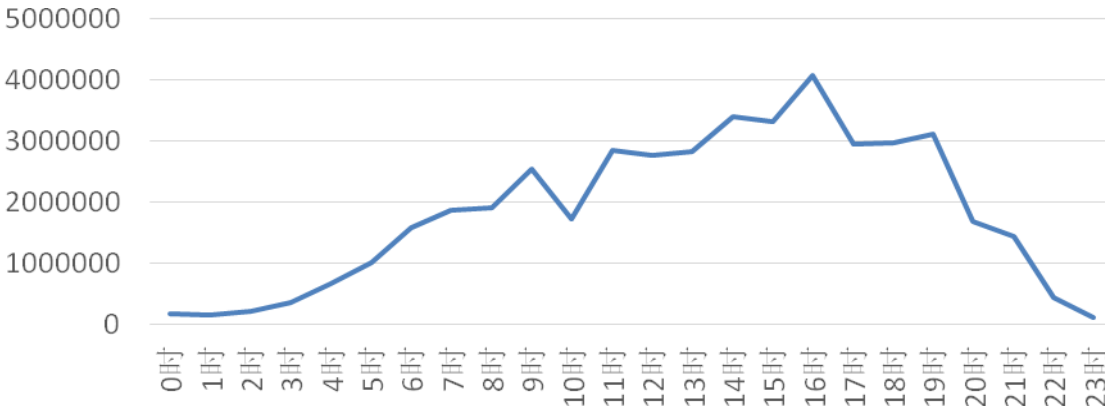


图 8 车站每小时总客流量折线统计图

图 8 是 2015-01-01 至 2016-03-20 期间车站每小时客流量折线统计图，该图以时间为横坐标，车站总客流量为纵坐标，研究一天内每个时间段的客流变化规律。

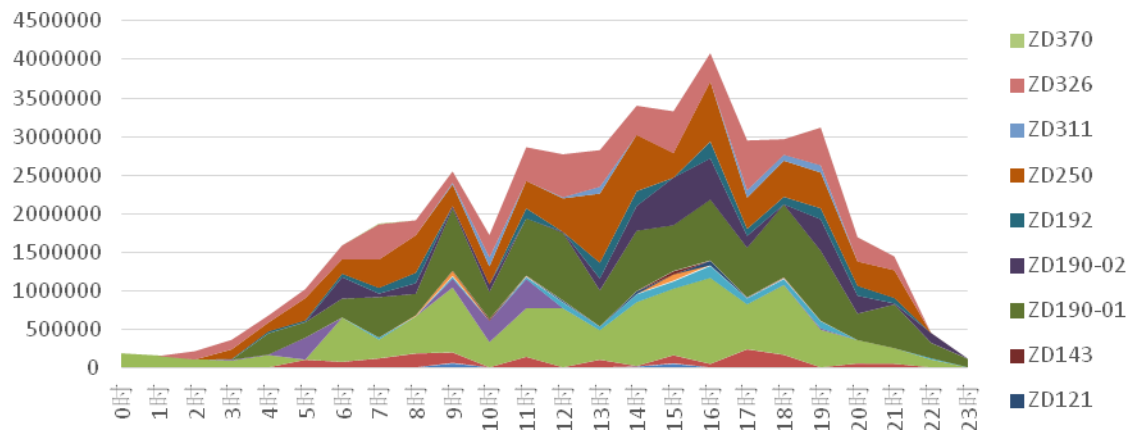


图 9 各车站每小时客流量面积图

图 9 是 2015-01-01 至 2016-03-20 期间各车站每小时客流量面积图，该图以时间为横坐标，总客流量为纵坐标，绘制出各站点客流量面积图，用于研究各站点一天内每个时间段的客流变化规律。

3. 客流量预测

对于未来铁路客流量预测，采用时间序列模型来进行预测，针对不同规律的序列分别构建平稳时间序列模型 ARMA 和非平稳时间序列模型 ARIMA。

3.1 ARMA 模型建模

根据气象因素与客流量相关性分析，引入气象状况相关变量，追加气象因素作为影响因素。通过对客流量序列进行预处理可以判定为平稳非白噪声序列，即可利用 ARMA 模型对该序列建模。

3.1.1 建模步骤

步骤 1、首先求出该序列的样本自相关系数 ACF 和样本偏自相关系数 PACF 值。

步骤 2、根据样本自相关系数和偏自相关系数性质，选择阶数适当的 ARMA (p, q) 模型进行拟合。

步骤 3、估计模型中未知参数的值。

步骤 4、检验模型的有效性。如果拟合模型未通过检验，回到步骤 2 重新选择模型拟合。

步骤 5、模型优化，充分考虑各种可能建立多个拟合模型，从所有模型中选择最优模型。

步骤 6、利用拟合模型，预测序列将来的走势。

3.2 ARIMA 模型建模

对于客流量序列通过预处理可以判定为平稳白噪声序列，就可以利用 ARIMA 模型对该序列建模。ARIMA 模型建模步骤与 ARMA 模型建模步骤类似。

3.3 建模与结果预测

通过处理站点到站点的客流量数据，一共得到 79 组站点到站点的下车人数，其中有 25 组站点到站点的下车人数由于数据量过少无法进行预测，所以仅预测 54 组站点到站点的下车人数。

通过对这个 54 组客流量序列分别建模，其中包含四种序列模型：ARMA 模型、ARIMA 模型。并且把未来两周拆分成两段，分别是 3 月 21 日至 4 月 1 日和清明两天（4 月 2 日和 4 月 3 日）分开进行建模。

构建时序模型采用 R 语言，需要加载程序包 zoo，

forecast, tseries。

说明：53 个站点的预测以及 R 语言源代码均整合在附件中，可供查看。

3.4 预测实例

此处引用预测 ZD022 站至 ZD190 站未来两周客流量作为示例。

首先对 ZD022 至 ZD190 的 15 个月的下车人数进行数据预处理，绘制时序图：

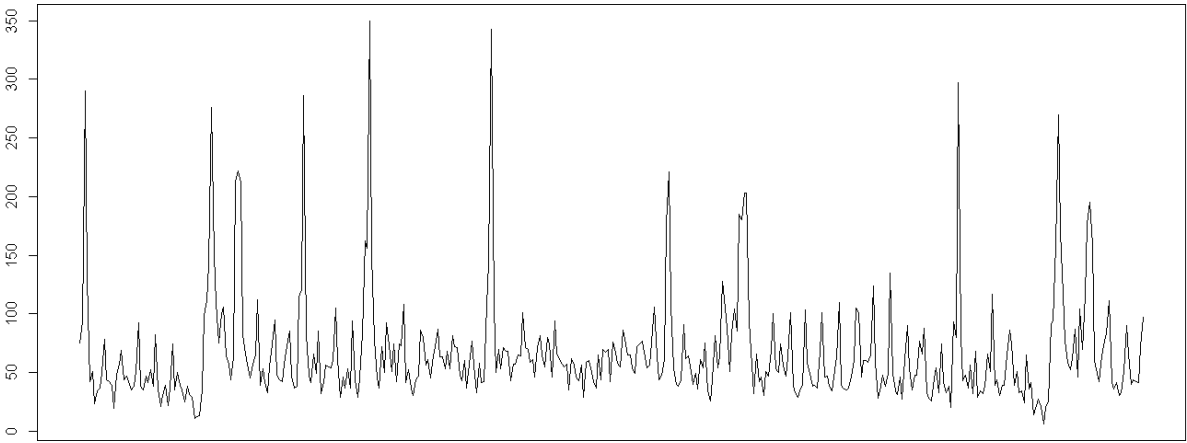


图 10 2015/1/1--2016/3/20 时间段 ZD022 至 ZD190 站点时序图

通过自动识别 auto.arima() 函数对数据进行分析，

```
Series: x
ARIMA(5,0,4) with non-zero mean

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ma1      ma2      ma3
      0.1011 -0.8941  0.3441 -0.2691 -0.0424  0.4383  1.2154  0.1802
s.e.    0.2277  0.1654  0.1969  0.1732  0.0954  0.2247  0.1524  0.2160
      ma4 intercept
      0.3667    66.7695
s.e.    0.1378     3.2735

sigma^2 estimated as 1456:  log likelihood=-2211.87
AIC=4445.74   AICc=4446.36   BIC=4490.64
```

图 11 数据分析结果

根据结果可以建立 ARIMA (5,0,4) 模型

$$x_t = 66.7695 + 0.1011x_{t-1} + (-0.8941)x_{t-2} + 0.3441x_{t-3} + (-0.2691)x_{t-4} + (-0.0424x_{t-5}) - 0.4383\varepsilon_{t-1} - 1.2154\varepsilon_{t-2} - 0.1802\varepsilon_{t-3} - 0.3667\varepsilon_{t-4}$$

利用所建立的模型，再使用 forecast () 函数对未来 12 天进行预测得出结果：

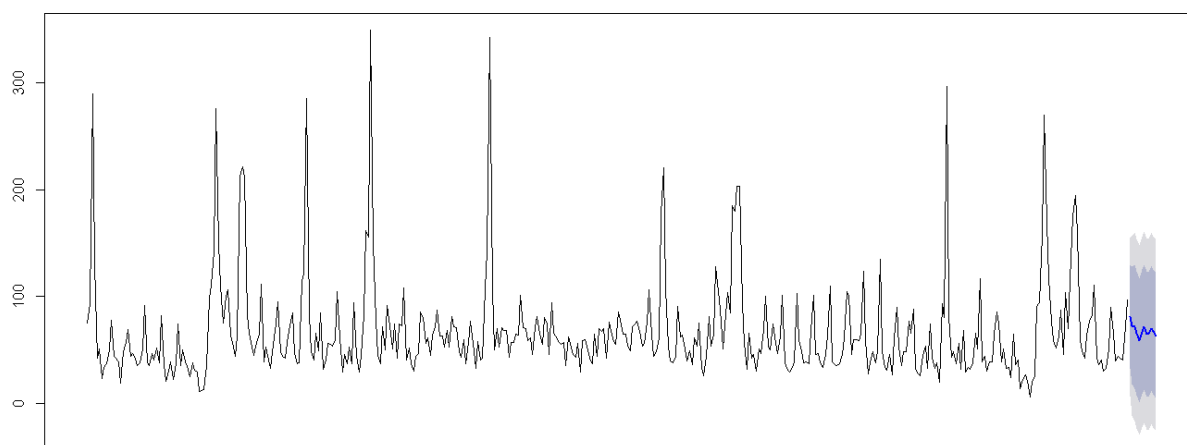


图 12 2015/1/1--2016/3/20 时间段 ZD022 至 ZD190 站

点时序预测图

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	81.11571	32.781240	129.4502	7.194515	155.0369
	73.01682	18.098063	127.9356	-10.974173	157.0078
	71.83301	13.987497	129.6785	-16.634070	160.3001
	65.38071	7.406300	123.3551	-23.283497	154.0449
	59.10773	1.092910	117.1225	-29.618280	147.8337
	66.68873	8.311952	125.0655	-22.590845	155.9683
	71.50622	13.031296	129.9811	-17.923458	160.9359
	64.84338	6.296328	123.3904	-24.696608	154.3834
	64.43267	5.866228	122.9991	-25.136975	154.0023
	70.23210	11.638860	128.8253	-19.378527	159.8427
	67.27506	8.678104	125.8720	-22.341250	156.8914
	63.23802	4.621061	121.8550	-26.408884	152.8849

图 13 2016/3/21--2016/4/1 预测结果

接着针对已知数据 15 个月中节假日下车人数进行预处理，对 4 月 2 日和 4 月 3 日建立相关时间序列模型进行相关预测。

```

Series: x
ARIMA(0,0,0) with non-zero mean

Coefficients:
            intercept
            139.5952
s.e.        13.9011

sigma^2 estimated as 8314:  log likelihood=-248.63
AIC=501.26   AICc=501.57   BIC=504.73

```

图 14 数据分析结果

通过自动识别 `auto.arima()` 函数对数据进行分析，建立 ARIMA (0,0,0) 模型，利用 `forecast()` 函数对 4 月 2 日和 4 月 3 日进行预测得出结果

```

Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
      139.5952  24.14138  255.0491  -36.9762  316.1667
      139.5952  24.14138  255.0491  -36.9762  316.1667
      .      .

```

图 15 4 月 2、3 日下车人数预测结果

由以上过程即可完成 ZD022 站点至 ZD190 站点对未来两周的预测。通过对比实际数据检验了模型精度，表明了该模型具有较高的精度和良好的适用性。

四、总结

1. 数据分析结果

1.1 天气相关性分析

从表 2 各城市的客流与天气参数值相关性系数中可得如下分析。

管内各城市的天气参数值与客流量都呈不同程度的负

相关。ZD111、ZD250、ZD326 这三个城市天气参数值与客流量成低度负相关，ZD190 市天气参数值与客流量呈中度负相关。

说明 ZD111、ZD250、ZD326 这三个城市的客流量受天气因素的影响不大，不影响客流规律的分析。ZD190 市的客流量受到天气因素一定程度的影响，分析客流规律时需要对极端天气进行误差处理。

1.2 各车站客流量分析

1.2.1 铁路客流量波动变化规律

通过图 5 分析可知，铁路客流量主要体现在按月波动方面，从月波动性变化情况分析可得其中规律性：2 ~ 3 、7 ~ 8 月份的客流量达到高峰，5 月初、10 月初客流量较大，而其它月份的客流量较小，其中 11、12 月为最低。

1.2.2 各阶段客流量突增原因分析

1.2.2.1 春运

春节前后，各小学、中学、大学以及大中专院校放假和开学，民工返乡和返城，产生大量学生流、民工流。此外，出差、经商、探亲等客流也大量增加，这些客流同时交织在一起，使客流量达到第一个高峰期。

1.2.2.2 暑运

各小学、中学、大学以及大中专院校的放假、开学形成大量的学生流、家长流，同时旅游旺季的到来使旅游流大幅

度增加；在此期间，农业生产进入农闲时期，使农民流、个体商贩流急剧上升。这些因素使客流量达到第二个高峰期。

1.2.2.3 “五一”和“十一”节日

随着经济发展和生活水平的提高，旅游业持续发展，使客运量快速增长。特别是假日的旅游热潮，极大地促进铁路客运量的增加。

1.2.3 统计图、表格的相关分析

从客流量折线统计图（图 5）中可以看出，ZD190-01 在全年中客流量都稳居高位，特别是暑假期间客流量明显增多，且在法定节假日时急剧上升。作为当地铁路线的最后一站，此站点客流量远远大于其他站点的客流量，对站点的承载力要求较高。此站点在当地应是一个具有相当大影响力的站点，而当地存在其他客流量小的站点，可以做适当调整，以配合该站缓解客流压力。

ZD143、ZD311 等站点在全年客流量都偏少，说明这些站点不处于城市中心地区，不作为当地主要车站，途径列车较少。此类车站使用潜能较大，可参照当地城市规划以及站点周边交通发展情况，增加停靠列车数量，缓解当地主要车站的客流压力，或者加开城际列车，发展城市快捷交通圈，以提高资源利用率。

由客流量相关性系数表（表 3）中可以看出，大部分车站之间的客流量呈中度正相关。说明管内车站的客流量规律

有一定的相同性，当管内某一个车站出现大幅度客流变化的时候，应通知其他车站做好相关工作，防止客流压力突增带来的相关问题。

ZD111-01、ZD250、ZD326 这三个车站两两之间的相关系数大于 0.95，说明这三个车站的客流日客流量曲线呈高度正相关。说明这三个站在管内铁路线上的地位相当，当一个车站出现客流变化时，另外两个车站很有可能出现相同情况。为此这三个车站可以形成协议，讨论出共同的方案，保证各站能在最短时间预防客流量的变化带来的问题。

1.3 各车站上下车人数规律分析

从各车站上车人数折线统计图（图 6）中可以看出，ZD190-01 等站点 7 月到 10 月上车人数普遍较高，说明此车站流出人口较多，是人流高峰期，可能是旅游专区，预留车票数量可以适当增加。

在各车站下车人数折线统计图（图 7）中我们可以明显看出，ZD111-01、ZD190-01 等站点对客流有较大吸引力度。

同时我们也可以得出推论：这些站点的流动人口数、城市化程度以及经济发展水平相对比其他站点高。

ZD143、ZD311 等站点在全年上下车人数都偏少，说明此站点客流量不高，车站资源利用率不高。

对比两张图可以发现，管内车站的下车人数明显大于上车人数。可以发现客流量有一部分是从管外车站流动到管内

车站。

管内各车站的上下车人数的相关性系数明显低于上车量。说明各站点间的客流流动不平均，尽管各车站客流量有相同规律，但是各站还是需要单独拿出方案应对乘车人流与抵达人流的变化。

1.4 车站每小时总客流量分析

由车站每小时总客流量折线统计图（图 8）按照时间序列分析一天中管内 15 个车站总客流变化趋势，反映出一天内的客流量显示出较强的不均衡性。本折线统计图还可以反映出一定的规律性：凌晨时间客流量较低，各个站点上午的高峰小时并不明显，从 6:00-7:00 开始客流量逐渐增多，16:00 到达峰值，之后缓慢下降。

人口流动在白天相对频繁，大多数旅客选择白天搭乘列车，具有多方面原因，分析如下：

不同地区经济发展水平不同，存在梯度性差异，因此不同地区的旅客总体消费水平也存在着较大差异。一天内管内所有车站总客流变化主要由各个人口密度较大、经济水平较高的地区所决定。平日里的客流以工人、企业人员和学生为主，主要出行目的是公务、会议、商务、务工、学习培训、探亲、旅游休闲等，都聚集在白天几个主要时间段。在春运、五一、十一等客运旺季时，客流主要还是学生流、探亲流和商务流等，旅客基本都会选择白天出行。充分体现了如今人

们生活节奏快，对工作和旅行的需求。

2. 建议

(1) 根据月客流量的波动变化，在春运、暑运、“五一”和“十一”客流高峰时，需要提前作好铁路运输组织工作，以迎接客流高峰的到来，避免安全事故的发生。当客流量处于低谷时，应调整虚糜列车编组，进行车辆整备、检修以及人员培训等工作。

(2) 对于 ZD143 等小站点，可减少车次以及合理安排车辆配置。

(3) 列车通过停站方式输送旅客，停站数目越多越有利于旅客的乘降，方便旅客出行。但列车中途停站过多会降低旅行速度，所以列车停站的设置既要考虑旅客出行的方便性，又不能过多地降低列车旅行速度，从而造成旅客旅行时间的增加。在此给出两个停站方案（适用于高速列车）：

- 仅在指定的个别客运站始发、终到，且在运行径路上除始发、终到站以外的任何客运站都不办理客运作业。
- 在选定的客运站上办理客运作业。

(4) 合理公道的票价是旅客选择出行方式的一个重要因素。对于我国而言，铁路票价对旅客的影响具有一定作用。消费水平不同的旅客对于“合理的票价”有着不同的衡量标准。建议铁路部门可以更多考虑票价的合理性和旅客的承受能力来确定灵活的、动态的票价。

(5) 对于客座率超过 100%的车次，说明需要售出站票。建议提高发车密度，以提高铁路客运部门提供给旅客的运输服务产品的数量水平。

(6) 客运站选址：

- 城市内的铁路客运站选址要考虑乘客从出发点至客运站站址的出行距离、出行时间最短，出行费用最省，方便旅客乘车为目的，减少和方便旅客换乘，尽量缩短旅客运输距离。
- 客流需求会产生市场效益，但对于客运站选址而言，应考虑客运站选址方案和运营费用的相关性，采取经济最优方案。
- 铁路客运站与城市相容，引导推动城市交通发展。

3. 结论

该铁路沿线各站的客流有不同特点，在一天内有各自明显的出行规律，大部分站点在节假日期间都会形成一定的高峰时段，旅客流动相对频繁。

基于所构建的铁路客流量时间序列 ARMA 和 ARIMA 模型，预测了 2016 年 3 月 21 日至 2016 年 4 月 3 日各站点客流量，并通过对比实际数据检验了模型精度，能够得到较为准确的客流量预测值，基本满足对客流量的把握需求。