

## 基于航班延误预测的多天气模式下时刻协调参数剖面研究

高伟<sup>1</sup>, 逢丁荧<sup>1,2</sup>

(1. 中国民航大学空中交通管理学院, 天津 300300; 2. 中国民用航空华北地区空中交通管理局终端管制室, 北京 100621)

**摘要:** 构造符合正常性期望的繁忙机场 18~24 h 合理时刻协调参数剖面用于公布容量和时刻换季是尚未解决的民航时刻管理难题。本文使用机场航班和天气历史数据, 通过 K-means 聚类及偏最小二乘回归建立时刻结构的回归预测模型, 利用集成学习预测航班延误水平。结果表明: 随机森林在回归和预测方面都呈现较好的效果, 并能结合航班延误预测得到时刻协调参数剖面的上限与下限作为时刻协调参数区间; 将结果进行仿真验证, 使得在区间内的航班架次安排满足小于 15 min 的平均延误时间水平, 最终给出建议的时刻协调参数。本文可为相关部门的时刻管理, 不同战略战术时期的空中交通流量管理, 以及机场和航空公司评估延误风险、调整时刻安排、配置运力 and 保障资源, 提供精细化的辅助决策信息。

**关键词:** 空中交通; 多天气模式; 航班时刻; 航班延误; 集成学习

**中图分类号:** V352 **文献标志码:** A **文章编号:** 1674-5590(2025)06-0024-07

### Research on coordinated slot parameter profiles based on flight delay prediction under multiple weather patterns

GAO Wei<sup>1</sup>, PANG Dingying<sup>1,2</sup>

(1. College of Air Traffic Management, CAUC, Tianjin 300300, China; 2. Approach Control Unit, North China Regional Air Traffic Management Bureau of CAAC, Beijing 100621, China)

**Abstract:** Constructing a reasonable 18—24 h coordinated slot parameter profile for busy airports that aligns with normality expectations, used for publishing capacity and facilitating seasonal slot adjustments, remains an unresolved issue in aviation slot management. This paper develops a regression prediction model for slot structure using historical flight and weather data, employing K-means clustering and partial least squares regression, while applying ensemble learning to forecast flight delay levels. The results show that random forests exhibit good performance in both regression and prediction, and can combine flight delay predictions to obtain the upper and lower limits of the coordinated slot parameter profile as the slot coordination parameter interval. The results are verified through simulation, ensuring that the flight schedule within this interval maintains an average delay level of less than 15 min, leading to the proposal of recommended slot coordination parameters. This paper provides refined decision-support information for relevant departments in slot management and air traffic flow management during different strategic and tactical periods, as well as for airports and airlines in evaluating delay risks, adjusting slot arrangements, allocating capacity, and planning support resources.

**Key words:** air traffic; multiple weather patterns; flight slots; flight delay; ensemble learning

航班时刻是“航空器在指定日期和时间, 为抵离某个机场而使用相关基础设施与服务的权利”<sup>[1]</sup>, 是民航的生产资料, 是航企运营的根本和供不应求的公共资源。Khan<sup>[2]</sup>研究了如何使用机场时刻协调参数作为繁忙机场平衡需求与供给的机制。时刻协调参数规定了机场在单位时间(1 h 或 15 min)内可供机场调整航班起降的数量, 实际上设置了机场可分配时刻总数的

上限, 是供给侧保障能力的反映<sup>[3]</sup>。在向智能化行业转型过程中, 中国民航注重时刻资源的精细化管理和提质增效, 主协调机场和辅协调机场等繁忙机场的时刻协调参数(亦称作时刻容量或公布容量)不再执行单值化管理, 转而按全天各小时公布<sup>[3]</sup>。即便如此, 形式上满足中国民用航空局(简称局方)对时刻协调参数约束的航班时刻表, 由于各时段时刻协调参数之和远

超出机场的日航班保障能力,蕴含着航班时刻的结构原生延误,这成为机场航班时刻安排实践和正常性管理中难以管控的隐性因素。批复时刻表的时间和架次与实际执行的时刻结构间仍存在差别,严重影响空中交通流量战略管理和战术控制的决策<sup>[4-5]</sup>。

公布容量需考虑容量利用率和服务质量(航班延误或准点率)之间的均衡。航班延误预测的研究多以既定时刻协调参数剖面结构为前提预测航班服务水平,而以满足服务水平为前提设置合理的航班时刻协调参数剖面结构则构成其反问题<sup>[3]</sup>。

时刻协调参数剖面结构既是市场需求变化的反映,也要满足运行效率约束。单机场时刻分配本质上属于资源受限的分配问题,目前主要由局方依靠运筹学理论方法配置,单机场稳健性公布容量确定方法仍是一个尚未解决的难题<sup>[3]</sup>。

Gillen 等<sup>[6]</sup>强调时刻协调参数与多方政策之间的相互依赖性。时刻协调参数的制定仍是迄今尚未解决的民航时刻管理难题<sup>[7]</sup>。为保障航班运行效率,国际实践中常使用防火槽(fire break)<sup>[8]</sup>在 2 个连续的繁忙时段后强制降低 1~2 架次,但该方法因缺乏理论支持,仅靠试错手段,实际效果不佳。

由于各时段时刻协调参数之间的复杂牵制作用,难以用解析方法解决满足效率目标的时刻协调参数结构问题,本文将时刻协调参数剖面看作一个整体,通过学习航班运行历史数据,结合天气影响因素,建立航班延误与时刻协调参数剖面之间的内在联系。在运行环境稳定时,从运行效率角度给出由航班延误预测驱动的协调参数剖面分布区间,并给出合理的期望剖面结构。

## 1 数据构建与处理

### 1.1 基础数据采集

航班延误与运营商、航线、中转衔接、时刻时段、空域流量及天气等因素有关。选取机场航班实际运行数据的 11 个有效特征项如表 1 所示,包括航班号、航空公司、前站起飞时间、计划起飞时间、实际起飞时间、计划落地时间、实际落地时间、航线、出发机场(航班起飞的机场)、落地机场及机型。选取的特征项与航班延误密切相关,航班时间信息直接影响到航班的延误,航班基本信息用于得知具体延误航班,航班路径信息用于根据航路拥堵情况预测延误。天气数据来源于羲和能源气象大数据平台,选取机场历史天气数据,每条天气数据包括机场、时间、大气温度、气象站

水平气压、平均海平面气压、气压变化、相对湿度、风向、风速、最大阵风、能见度、露点温度、降水量和天空状况等 14 个特征项。其中部分天气数据如表 2 所示。

表 1 航班实际运行数据的有效特征项

Tab.1 Informative features derived from actual flight operation data

编号	特征项	示例
1	航班号	768
2	航空公司	CES
3	前站起飞时间	2018-08-04 07:40:00
4	计划起飞时间	2018-08-04 10:40:00
5	实际起飞时间	2018-08-04 10:40:00
6	计划落地时间	2018-08-04 14:00:00
7	实际落地时间	2018-08-04 14:00:00
8	航线	TPE—HRB
9	出发机场	ZBAA
10	落地机场	ZYHB
11	机型	A320

表 2 天气数据集有效特征项

Tab.2 Informative features in weather datasets

编号	特征名称	特征说明	示例
1	State	机场	ZYHB
2	Date	时间	23:00
3	AtmosphericTemperature	大气温度/°C	-11.7
4	StationPressure	气象站水平气压/hPa	763.0
5	SeaLevelPressure	平均海平面气压/hPa	775.0
6	PressureChange	气压变化/hPa	-0.8
7	RelativeHumidity	相对湿度/%	71
8	WindDirection	风向/(°)	70
9	WindSpeed	风速/(m·s <sup>-1</sup> )	4
10	MaximumGusts	最大阵风/(m·s <sup>-1</sup> )	7
11	Visibility	能见度/km	12.0
12	DewPointTemprature	露点温度/°C	-15.9
13	Precipitation	降水量/mm	0
14	SkyC	天空状况	OVC

### 1.2 数据处理

对机场航班实际运行数据和天气信息进行数据清洗,删除多余信息,补齐缺失信息。由于各个特征向量的维度不同,很可能对实验结果造成影响,因此采用 min-max 标准化的方式对数据进行归一化处理,将原始数据  $z$  映射到区间 $[0, 1]$ 中,得到标准化后的数据

$$z' = \frac{z - z_1}{z_2 - z_1} \quad (1)$$

式中: $z_1$  表示原始数据中的最小值; $z_2$  表示原始数据中的最大值。

## 2 实验与分析

### 2.1 航班时刻协调参数剖面构造

选取哈尔滨太平国际机场(简称哈尔滨机场)

2018 年 4 月 9 日—2023 年 4 月 4 日历史数据进行算例分析,在此期间,哈尔滨机场的运行环境总体稳定。通过历史数据的学习,建立合理的时刻协调参数剖面结构。以小时为单位,哈尔滨机场 2018 年 4 月 9 日各时间段的进离港航班架次示例如表 3 所示。

表 3 各时间段进离港航班架次示例

Tab.3 Example of arrival and departure flights by time period

计划时段	离港航班	进港航班
00:00:00	1	0
01:00:00	0	0
02:00:00	0	0
03:00:00	0	0
04:00:00	2	6
05:00:00	1	3

为研究机场进离港航班架次时刻结构分布,以一天的时刻协调参数剖面作为聚类变量,采用 K-means 聚类算法对进离港航班架次聚类,计算误差平方和 (SSE, sum of squared errors) 值,即

$$V_{SSE} = \sum_{j=1}^K (c_j - m_j)^2 \quad (2)$$

式中:  $V_{SSE}$  表示误差平方和,用于衡量聚类的紧凑程度;  $K$  表示聚类数;  $m_j$  表示第  $j$  个簇中的数据点;  $c_j$  表示第  $j$  个簇的质心。

聚类数  $K$  与 SSE 值的折线图如图 1 所示。

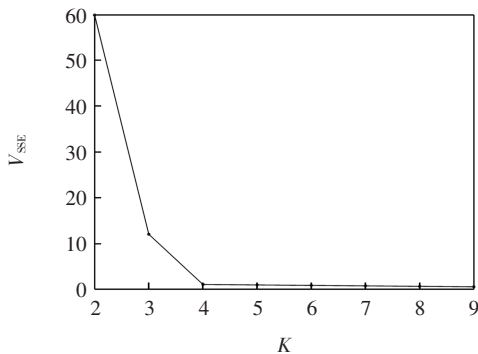


图 1 K-means 聚类数与 SSE 值折线图

Fig.1 Line plot of  $V_{SSE}$  vs. number of K-means clusters

根据“肘部法则”,由图 1 可知,聚类数  $K = 4$  时,可以在一定程度上保持聚类效果和模型复杂度的平衡,因此得到 4 种时刻协调参数剖面结构,并取其平均值,如图 2 所示。

由图 2 可知,聚类 3 整体架次偏高,聚类 2 整体架次偏低,聚类 1 和聚类 4 整体架次居中。聚类 1 代表哈尔滨机场冬春航季,聚类 2 代表哈尔滨机场淡季,聚类 3 代表哈尔滨机场旺季,聚类 4 代表哈尔滨机场夏秋航季。

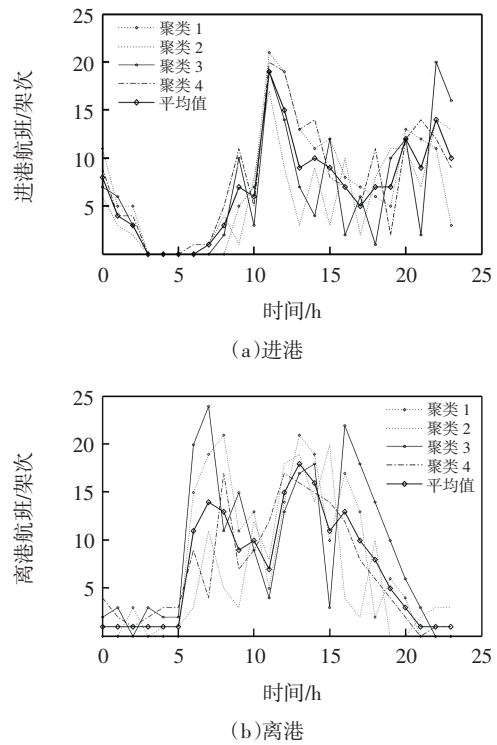


图 2 一天的时刻协调参数剖面结构聚类结果

Fig.2 Clustering results of daily coordinated slot parameter profile structures

## 2.2 基于随机森林的航班延误预测模型

随机森林(RF, random forest)属于并行集成学习中 Bagging(bootstrap aggregating)算法的一种扩展变体,其在以决策树为基学习器构建 Bagging 集成的基础上,进一步在决策树的训练过程中引入随机属性选择<sup>[9-18]</sup>。

将表 1 中的特征项作为运行数据特征变量,表 2 中的特征项作为天气数据特征变量,建立基于 RF 的航班延误预测模型,对航班是否延误进行预测。

### 2.2.1 分类模型

定义超参数搜索空间并构建 RF 分类模型。通过 Optuna 进行超参数优化,找到最佳的超参数组合,即决策树数量为 100、决策树最大深度为 5、叶节点的最小样本数为 4、节点划分的最小节点数为 3。

对模型进行评估,绘制受试者工作特征(ROC, receiver operating characteristic)曲线。ROC 曲线是根据一系列不同的二分类方式(分界值或决定阈),以真阳性率(灵敏度)为纵坐标,假阳性率(1-特异度)为横坐标绘制的曲线,可以直观地观察到分析方法的准确性。

AUC(area under the ROC curve)值是 ROC 曲线下方的面积,在  $AUC \geq 0.5$  的情况下,其值越接近于 1,说明模型效果越好。模型在  $0.5 \leq AUC \leq 0.7$  时准确性较

低; 在  $0.7 < AUC \leq 0.9$  时有一定准确性; 在  $AUC > 0.9$  时有较高准确性。AUC 值计算公式表示为

$$S_{AUC} = \frac{1}{2} \sum_{i=1}^{r-1} (x_{i+1} - x_i)(y_i + y_{i+1}) \quad (3)$$

式中:  $r$  表示模型训练集的数量;  $x_i, y_i$  分别表示  $i$  点的横、纵坐标值。根据计算得出的训练集和测试集上的 AUC 值绘制 ROC 曲线, 如图 3 和图 4 所示。

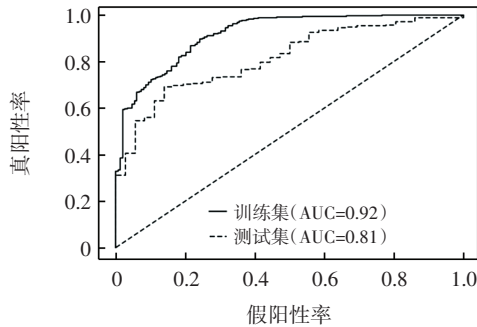


图 3 随机森林 ROC 曲线图

Fig.3 ROC curve of RF

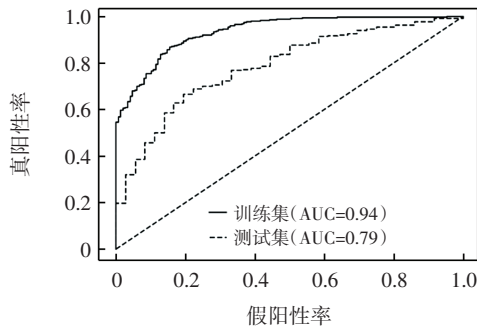


图 4 支持向量机 ROC 曲线图

Fig.4 ROC curve of the support vector machine

由图 3 可知, RF 分类模型在对新数据进行预测方面能提供更好的泛化能力。将航班运行数据作为测试集输入已经建立的 RF 分类模型和支持向量机 (SVM, support vector machine) 分类模型中进行是否延误预测, 结果显示 RF 分类模型的准确率为 98%, SVM 分类模型的准确率为 96%。因此, 选取 RF 分类模型作为数据训练的分类预测模型。

### 2.2.2 回归模型

将上述数据划分为测试集和训练集, 其中测试集的比例为 20%。建立 RF 回归模型完成回归预测, 并尝试不同的决策树数量和决策树最大深度。

经迭代得到构建 RF 回归模型的参数有: 决策树数量为 200、每个决策树的最大深度为 12、随机种子数为 420。

航班延误是指航班实际起飞或落地时间(航班实际出港撤轮挡时间或进港挡轮挡时间)比计划起飞或

落地时间(航班时刻表上的时间)延误 15 min 以上或航班取消的情况<sup>[7]</sup>。通过 RF 回归模型预测每个航班的延误时间, 筛选出延误时间大于 15 min 的航班作为延误航班。表 4 为部分预测值和真实值的结果示例。

表 4 部分预测值和真实值比较

Tab.4 Comparison of partial predicted values and actual values

平均进港延误时间		平均离港延误时间	
真实	预测	真实	预测
-8.0	-24.7	-4.0	-11.9
94.0	88.5	16.0	26.9
-7.2	-6.4	-3.0	-3.8
23.0	48.2	15.0	8.2
-21.0	-9.3	119.0	111.5

将 RF 回归模型与 SVM 回归模型进行比较, 使用相同的训练集和测试集, 其中,  $R^2$  表示决定系数, MAPE 表示平均绝对百分比误差。模型比较结果如表 5 所示。

表 5 模型比较

Tab.5 Comparison of models

模型	$R^2$	MAPE
RF	0.813	14.243
SVM	0.742	18.745

由表 5 的结果可以看出, RF 回归模型预测的效果比 SVM 回归模型更好。因此, 选用 RF 回归模型进行延误预测。

## 2.3 航班延误预测驱动的时刻协调参数剖面构造

### 2.3.1 偏最小二乘回归预测模型的建立

将 2.2.2 节预测的延误时间数据进行整理并筛选出延误时间大于 15 min 的相关数据即为延误航班, 统计每小时延误航班架次并用 K-means 聚类算法对一天内各个小时时段的延误航班数进行聚类(同 2.1 节), 得到 24 h 进、离港延误航班架次的 4 种时刻结构, 如图 5 所示。

由图 5 可以看出, 4 种航班时刻协调参数剖面总体变化趋势相同, 仅有总量和延误架次的差异。取航班历史数据平均值作为基准, 得到哈尔滨机场平均航班时刻协调参数剖面, 如图 6 所示。

从图 6 可以看出, 进港航班和进港延误航班架次走势大致相同, 离港航班和离港延误航班架次走势大致相同。受跑道和空域环境约束, 离港航班和进港航班之间相互影响制约, 在进港航班数量高的时候, 相对应的离港航班数量偏低, 二者具有高度相关性。

偏最小二乘回归 (PLSR, partial least squares regression) 是一种适用于处理因变量和自变量之间存在

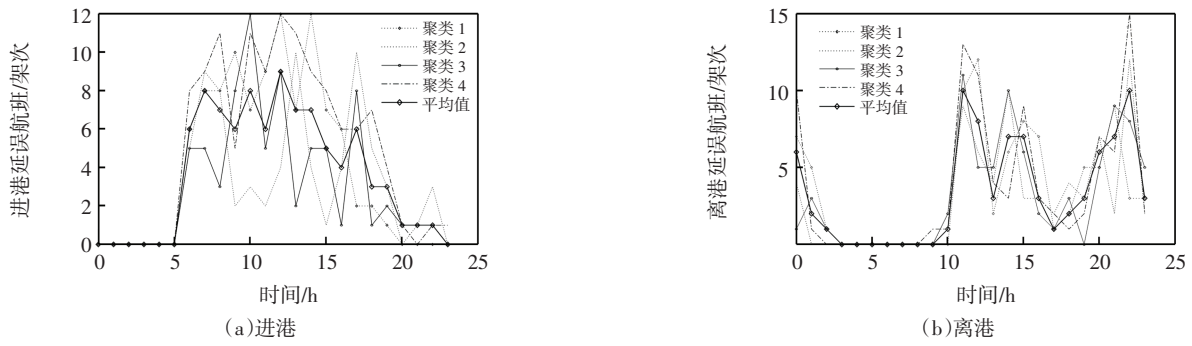


图 5 进、离港延误航班架次时刻结构

Fig.5 Slot structure of delayed arrival and departure flights

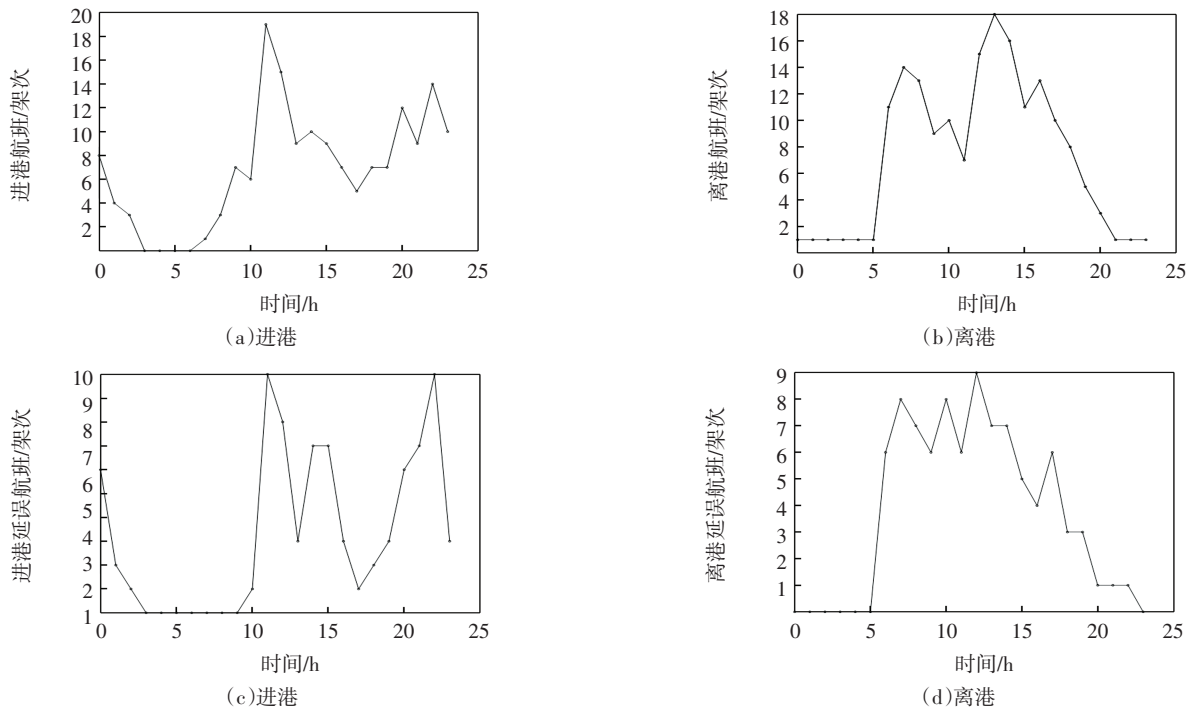


图 6 平均航班时刻结构

Fig.6 Average flight slot structure

多重共线性的多变量线性回归分析方法。

对进离港航班及其延误航班架次进行 PLSR, 同时考虑自变量与自变量之间的相关性以及自变量与因变量之间的回归关系, 得到 PLSR 预测模型参数, 如表 6 所示, 其中:  $t$  值用于检验回归系数是否显著不为 0, 即回归系数与标准误差的比值;  $p$  值为  $t$  检验的概率值。

表 6 模型回归系数检验

Tab.6 Regression coefficient test of the model

因变量 $Y$	自变量 $X$	回归系数	标准误差	$t$ 值	$p$ 值
离港航班架次	进港航班架次	-0.159	0.227	-0.700	0.489
	离港延误航班架次	1.710	0.142	12.040	0
进港延误航班架次		0.150	0.339	0.442	0.663

可得因变量  $Y$  和所有自变量  $X$  的回归关系式为

$$Y = -0.158X_1 + 0.321X_2 + 0.879X_3 \quad (4)$$

式中:  $Y$  为离港航班架次;  $X_1$  为进港航班架次;  $X_2$  为离港延误航班架次;  $X_3$  为进港延误航班架次。

### 2.3.2 满足服务水平的航班时刻协调参数剖面构造

**第 1 步** 下载一个航季的天气数据, 输入基于 RF 的航班延误预测模型中进行训练, 得到每天各时段的航班延误时间, 筛选出延误航班并统计每个时段的进离港延误航班架次。

**第 2 步** 对天气数据进行降维和聚类。利用主成分分析法降维并绘制权重占比折线图, 如图 7 所示。绘制  $V_{SSE}$  与聚类数  $K$  的关系图, 如图 8 所示。

由图 7 观察主成分的权重比例, 保留前几个权重比例较高的主成分作为降维后的特征。

根据图 8 选择  $V_{SSE}$  增长开始变缓之前的点来确定聚类中心的个数。

结合图 7 和图 8, 选择前 6 个主成分及 5 个聚类

中心对天气数据进行 K-means 聚类。图 9 为基于 6 个主成分的天气数据聚类结果可视化(聚为 5 个类别), 以方差贡献最高的主成分 1、主成分 2 为坐标轴, 呈现了 5 类天气在主成分空间中的分类簇。

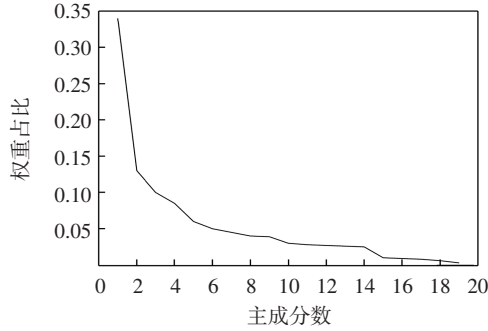


图 7 权重占比折线图

Fig.7 Line chart of weight proportion

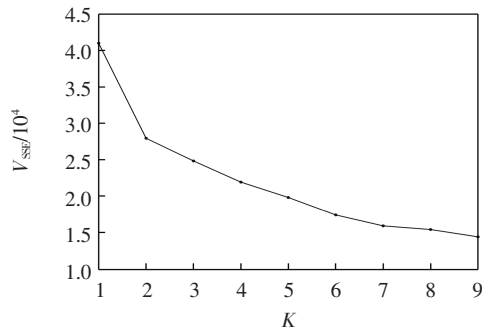


图 8  $V_{SSE}$  与聚类数  $K$  的关系图

Fig.8 Relationship between  $V_{SSE}$  and the number of clusters  $K$

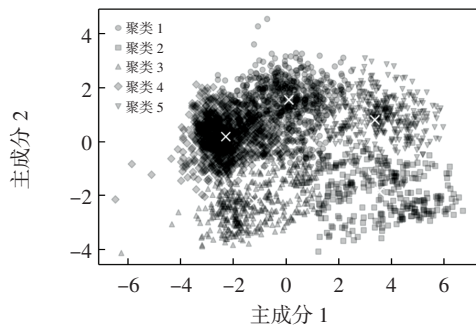
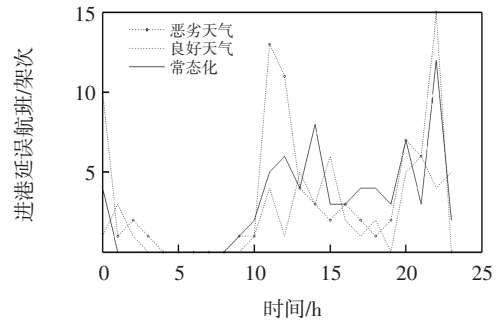


图 9 天气聚类结果图

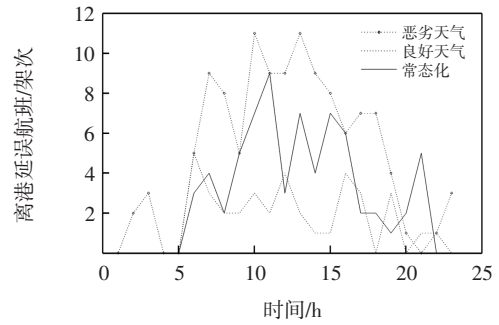
Fig.9 Clustering results of weather data

根据图 9 分别选择恶劣天气和良好天气及其对应的进、离港延误航班架次并求其平均值, 其中恶劣天气代表每小时的降水量超过 3 mm, 风速大于 17.2 m/s; 良好天气代表每小时的降水量为 0 mm, 风速小于 5 m/s。其他天气特征对应的进离港延误航班架次的平均值作为常态化时刻结构, 结果如图 10 所示。

第 3 步 将上述结果及计划进港航班架次或计划离港航班架次输入 PLSR 预测模型中, 得到不同时间特征下的进、离港航班架次时刻结构, 如图 11 所示。



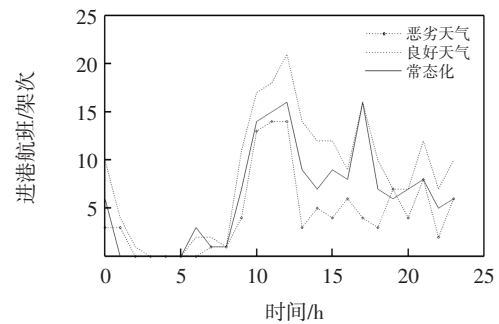
(a)进港



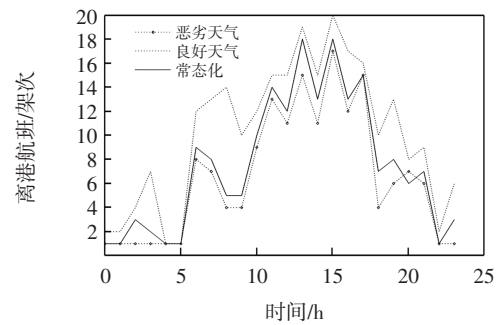
(b)离港

图 10 不同天气特征下进、离港延误航班架次时刻结构

Fig.10 Slot structure of delayed arrival and departure flights under different weather characteristics



(a)进港



(b)离港

图 11 不同天气特征下进、离港航班架次时刻结构

Fig.11 Slot structure of arrival and departure flights under different weather characteristics

图 11 的时刻结构中每个时段时刻参数波动范

围对应 3 条曲线的上下限之间。其中：剖面下限曲线代表恶劣天气；上限曲线代表良好天气；上下限曲线之间代表常态化天气下的时刻结构。在恶劣天气下进港航班架次安排少于良好天气，在一定程度上可以减少延误，在区间内的航班架次安排满足航班正常性水平及市场需求，为机场制定计划时刻结构提供了帮助。

针对一个航季中可能出现的各天气模式计算航班时刻综合剖面，即每种天气模式出现的概率与其对应时刻协调参数剖面乘积之和。

**第 4 步 仿真验证。**根据第 3 步得到的期望剖面，利用蒙特卡洛仿真方法生成天气参数和航班计划，并通过空域与机场仿真工具进行仿真。对于每次仿真得到的航班计划，计算平均延误时间。根据仿真结果，判断平均延误时间是否满足 15 min 的要求。如果平均延误时间不满足要求，可以通过加密航班或者减少航班来调整剖面模式，重新进行仿真和评估，直至满足要求为止。根据模拟和仿真，得到建议的航班时刻协调参数如表 7 所示。

表 7 建议的航班时刻协调参数

Tab.7 Recommended flight slot coordination parameters

				架次			
时段	恶劣天气	良好天气	常态化	时段	恶劣天气	良好天气	常态化
00:00—00:59	4	12	7	12:00—12:59	25	32	31
01:00—01:59	4	6	5	13:00—13:59	18	31	24
02:00—02:59	1	5	2	14:00—14:59	16	27	25
03:00—03:59	1	7	2	15:00—15:59	21	32	26
04:00—04:59	1	4	1	16:00—16:59	18	26	24
05:00—05:59	1	2	0	17:00—17:59	19	32	20
06:00—06:59	8	14	9	18:00—18:59	7	20	19
07:00—07:59	8	15	9	19:00—19:59	13	20	18
08:00—08:59	5	15	7	20:00—20:59	11	15	15
09:00—09:59	8	21	9	21:00—21:59	14	21	14
10:00—10:59	22	29	23	22:00—22:59	6	19	15
11:00—11:59	27	31	29	23:00—23:59	9	24	17

### 3 结语

时刻资源的合理分配是提高航空运输效率的关键，延误是重要的效率参数。从延误角度出发解决时刻协调参数问题需要首先建立延误与起降航班架次之间的联系。

相较于 SVM，基于 RF 分类与回归模型的航班延

误预测拟合度更高，准确率更好。回归模型建立了离港航班架次与进港航班架次及其延误航班架次之间的数学关系；同时通过聚类将复杂的天气划分为恶劣天气、良好天气和常态化天气，并对应生成 3 类天气下的延误航班架次剖面。在此基础上，利用回归模型分别得到时刻协调参数剖面的上限、下限和期望结构。最后通过模拟和仿真，将考虑天气的时刻协调参数从预战术层面转为战略层面，得到满足延误水平的建议时刻协调参数。

由于天气信息复杂，本文仅根据聚类结果大致区分出恶劣天气、良好天气和常态化天气，在后续相关研究中可以将常态化天气信息做更细化的分类与研究，更全面、合理地航班时刻协调参数结构构造提供完整的研究基础。

在运行环境稳定的前提下，未来可以根据天气聚类的细化给出更加精细的延误预测与时刻协调参数剖面构造，并能得到机场在不同航季的合理日航班架次水平。这一成果能够为局方的时刻管理，不同战略战术时期的空中交通流量管理，以及机场和航空公司评估延误风险、调整时刻安排、配置运力 and 保障资源提供精细化的辅助决策信息。

### 参考文献：

- [1] 中国民用航空局. 民航局关于印发《民航航班时刻管理办法》的通知[J]. 中华人民共和国国务院公报, 2018(19): 79-86.
- [2] KHAN A M. Airport slot controls[M]//Handbook of Transport Systems and Traffic Control. Bradford, West Yorkshire: Emerald Group Publishing Limited, 2001: 579-590.
- [3] 王艳军, 水笑雨, 王梦尹. 机场航班时刻资源管理研究进展[J]. 北京航空航天大学学报, 2024, 50(4): 1065-1076.
- [4] 禹常冠. 空中交通综合流量预测系统[J]. 民航管理, 2016(11): 52-57.
- [5] 杨波. 从同一航班号的时刻差异问题谈航班时刻协调参数管理[J]. 中国民用航空, 2019(1): 25-28.
- [6] GILLEN D, JACQUILLAT A, ODONI A R. Airport demand management: the operations research and economics perspectives and potential synergies[J]. Transportation Research Part A: Policy and Practice, 2016, 94: 495-513.
- [7] 交通运输部. 航班正常管理规定[J]. 中华人民共和国国务院公报, 2021(28): 22-31.
- [8] OFFERMAN H. Simulation to support the airport stakeholder decision-making process[J]. Air & Space Europe, 2001, 3(1/2): 60-67.
- [9] 周丰. 基于 K-means 聚类算法的航材故障预警方法研究[J]. 民航学报, 2022, 6(2): 90-92, 103.

(下转第 60 页)