【航空管理与安全】

基于灰色/回归＋马尔可夫模型的航班流量预测

赵江鸽，周洁敏

（南京航空航天大学 民航学院，江苏 南京 211100）

摘　要：合理准确的航班流量预测是制定相关管制方案和地面保障程序的前提。建立灰色/回归＋马尔可夫组合模型进行航班流量预测，并以机场实际航班流量数据为例进行模型验证。首先使用灰色模型和多项式回归对航班流量的训练样本进行趋势性预测，选取精度高的模型作为趋势模型。其次用马尔可夫模型对预测结果进行修正并比较修正前后平均相对误差，结果表明，修正后的误差从1.07％降低到0.64％。最后对测试样本进行验证，其相对误差在5％以内，验证了模型的可行性。此研究可为机场及时调整管制方案、提高运行效率、保障运行安全提供理论依据。

关键词：航班流量预测；灰色预测；多项式回归；马尔可夫链

中图分类号：V 355　　**文献标识码**：A　　DOI：10.13486/j.issn.2097-4973.2024.01.012

0　引言

准确、高效的航班流量预测技术，是实施科学的空中交通流量管理的基础。针对航班流量预测问题，国内外学者基于不同的模型算法开展了大量研究。向征等[1]通过航空器飞行计划信息和空域结构数据进行流量预测，可在一定程度上反应流量变化趋势；王兴隆等[2]提出一种基于机场相似日的粒子群优化支持向量机流量预测方法，该法预测稳定性较好；郭宪超[3]对不同阶段的历史ADS-B航迹数据进行聚类，并修正得到预测航迹，依此来进行飞行流量预测；闫少华等[4]使用门控循环单元神经网络模型和自回归移动平均模型的组合预测方法得到最终的预测结果；尚然然等[5]提出了一种结合聚类分析与LSTM-BP的终端区短期流量预测方法进行未来时段的终端区流量预测；KHAFIDLI等[6]提出了长短期记忆（LSTM）技术来预测印度尼西亚的航班流量，但在变量选取过程中仅考虑部分变量，未综合考虑所有影响因素；LU等[7]将贝叶斯网络应用于空中交通流量预测，可以有效预测未来时间段内各子域的飞机数量；朱承元等[8]使用了向量误差修正模型（VECM）与时间序列模型（ARIMA）进行航班流量预测计算，最终得出VECM比ARIMA的预测结果准确度更高，证明所提出的预测方法合理。以上模型算法都可取得较好的预测效果，但对于小样本数据存在过拟合现象，预测精度不高。对于小样本数据的预测问题，采用传统的回归拟合、灰色预测和马尔可夫模型的较多。商博明等[9]运用新陈代谢的思想建立了新维灰色理论模型，并用马尔可夫链理论优化预测值，结果表明该法预测精度高，效果较好；李剑等[10]对GNSS值进行预测，在改进灰色预测的基础上增加马尔可夫修正环节，提高了系统的稳定性；丁天明等[11]对港口货物吞吐量历年数据建立灰色GM(1,1)模型，并用马尔可夫链进行修正，最后用粒子群算法迭代寻优，结果表明该法预测值

收稿日期：2024-04-01

第一作者简介：赵江鸽（2000—　），女，山西临汾人，硕士研究生，主要从事交通运输规划与管理研究。
E-mail：zhaojiangge@nuaa.edu.cn

拟合程度更高。

以上研究采用灰色模型和马尔可夫模型进行预测，并在不同方面进行了优化，均得到较好的预测效果，航班流量与上述研究内容均具有多影响因素、小样本和随机性的特点，因此本文采用灰色马尔可夫组合模型进行预测。将预测分为两个阶段[12]：趋势性预测和随机性预测。在趋势性预测中加入了灰色预测模型和回归模型预测效果的比较，选取预测精度高的为最终趋势模型；然后用马尔可夫模型对趋势性预测结果进行修正，与航班流量随机性的特点相匹配。灰色/回归＋马尔可夫的组合模型的应用可以为较为准确的航班流量预测提供一种研究方法，为机场调整管制方案并保障飞行安全提供参考。

1 基于航班流量的预测模型

1.1 灰色预测模型

灰色预测模型适用于数据量少或数据质量较差，变化规律不明显的情况。对于机场航班流量，由于航班安排、航线、空域限制等因素，所运行的航班在一定程度上是固定的、有规律的，即灰色预测系统中的确定性信息。但一些突发情况，例如天气情况、人为因素、机械故障、疫情防控等因素，会对起飞和降落飞机航班次数产生影响，即灰色预测系统中的不确定性的信息。因此，在对航班流量进行预测时可选择使用灰色预测模型对原始数据的趋势特征进行分析。

根据时间序列，将航班流量的原始时间序列设为 $X^{(0)}=\{X^{(0)}(i), i=1,2,\cdots,n\}$。在灰色预测模型的实际应用过程中，需要检验该数列是否符合模型所需的光滑度条件，可通过计算参数来进行检验。当航班流量的原始数据无法达到光滑条件时，需对原始数据 $X^{(0)}$ 进行滑动平均处理，再建立灰色模型，形成滑动灰色模型。即对 $X^{(0)}$ 进行一次累加，生成一次累加序列：$X^{(1)}=\{X^{(1)}(k), k=1,2,\cdots,n\}$。其中 k 代表时间序列，即对航班流量按照时间顺序排序后赋予的序号，第一个数据即为 $k=1$，第二个数据即为 $k=2$，以此类推。

建立 GM(1,1)灰色预测模型

$$X^{(1)}(k)=\left(X^{(0)}(1)-\frac{u}{a}\right)\mathrm{e}^{-a(k-1)}+\frac{u}{a}。\tag{1}$$

式中：a 是发展灰数，u 是内生控制灰数；a，u 为需要根据实际数据计算所得出的参数，处于待定状态。当 a 的绝对值越大时，该模型的相对误差也就越大，通常来说有以下结论：当 $|a|\leqslant 0.3$ 时，适合中长期预测；当 $0.3<|a|\leqslant 0.5$ 时，适合短期预测；当 $0.5<|a|\leqslant 0.8$ 时，需格外注意；当 $0.8<|a|\leqslant 1$ 时，需要修正误差；当 $|a|>1$ 时，不宜使用该模型进行预测。

基于式(1)需要，根据实际航班流量数据计算得出 u、a，公式为 $\hat{\boldsymbol{a}}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_n=[a,u]^{\mathrm{T}}$。式中：

$$\boldsymbol{B}=\begin{bmatrix} -\frac{1}{2}(X^{(1)}(1)+X^{(1)}(2)) & 1 \\ -\frac{1}{2}(X^{(1)}(2)+X^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(X^{(1)}(n-1)+X^{(1)}(n)) & 1 \end{bmatrix}, \boldsymbol{Y}_{(n)}=(X^{(0)}(2),X^{(0)}(3),\cdots,X^{(0)}(n))^{\mathrm{T}}。$$

将计算得出的 u、a 代入灰色预测模型，根据时间顺序计算得出最后所要的结果。计算航班流量的预测值与真实值之间的相对误差并用百分数表示，可以更加有力地反映出预测数值的精确程度。

1.2 多项式回归模型

多项式回归模型是一种线性回归模型的扩展，是基于统计理论建立的最基本最常用的一类数据驱动模型，用于建立自变量和因变量之间的非线性关系。与简单线性回归模型不同的是，多项式回归模型允许自变量的多项式项存在，可以更灵活地适应数据的形态。根据原始数据趋势图可知，航班流量与时间之间

呈现非线性回归关系，但无法通过观察得出其适配于何种常见曲线，且航班流量数据中存在各种复杂的趋势形态，因此可采用多项式回归进行拟合。

设变量的 x,y 回归模型为 $y=\beta_0+\beta_1 x+\beta_2 x^2+\cdots+\beta_p x^p+\varepsilon$。式中：$y$ 是因变量即航班流量，x 是自变量即时间序列，$\beta_0,\beta_1,\cdots,\beta_p$ 是模型的系数；p 是多项式的阶数，决定了模型的复杂度；ε 是误差项，表示模型无法完全拟合的部分。

1.3　马尔可夫优化

马尔可夫过程是概率论研究中一种重要的随机过程，其状态在时间上具有马尔可夫性质，即未来处于何种状态只与当前状态有关，而与过去状态无关，这便是马尔可夫预测方法的一个最重要的核心。根据机场航班流量无后效性的特点，本文使用马尔可夫模型对趋势性预测模型的结果进行优化。

将所要预测的对象即航班流量按照预测的目的进行划分，本文根据历史数据的预测值与实际值之间的相对误差大小分为不同的状态进行标记，记为 E_i。计算不同状态之间的状态转移概率，$P(E_i \to E_j)$ 为由状态 E_i 转移到状态 E_j 的状态转移概率，将其表示为条件概率 $P(E_i/E_j)$，记作 P_{ij}。

由以上所得结果列出状态转移概率矩阵 $\boldsymbol{P}$，记作

$$\boldsymbol{P}=\begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & & \vdots \\ P_{n1} & \cdots & P_{nn} \end{bmatrix},$$

其中 P_{ij} 需要满足条件：

$$\begin{cases} 0 \leqslant P_{ij} \leqslant 1(i,j=1,2,\cdots,n), \\ \sum_{j=1}^{n} \boldsymbol{P}=1(i,j=1,2,\cdots.n)。 \end{cases} \quad (2)$$

式中计算矩阵可使用频率近似概率的思想。

计算经过 k 次转移后各月航班流量处于各个状态的概率即为状态概率 $\pi_j(k)$，将历史数据的最后数值作为初始状态($k=0$)，其中 $\sum_{j=1}^{N}\pi_j(k)=1$。从初始状态经过 k 次转移之后到达状态 E_j，这一个过程可以看作是初始状态经过 $k-1$ 次转移到达 $E_i(i=1,2,\cdots,n)$，再由 E_i 经过一次状态转移到达状态 E_j，则有公式

$$\pi_j(k)=\sum_{i=1}^{n}\pi_i(k-1)P_{ij}(j=1,2,\cdots,n)。\quad (3)$$

记行向量 $\boldsymbol{\pi}(k)=[\pi_1(k),\pi_2(k),\cdots,\pi_n(k)]$，由式(2)可得出状态概率的递推公式：

$$\begin{cases} \boldsymbol{\pi}(1)=\boldsymbol{\pi}(0)\boldsymbol{P}, \\ \boldsymbol{\pi}(2)=\boldsymbol{\pi}(1)\boldsymbol{P}=\boldsymbol{\pi}(0)\boldsymbol{P}^2, \\ \vdots \\ \boldsymbol{\pi}(k)=\boldsymbol{\pi}(k-1)=\boldsymbol{\pi}(0)\boldsymbol{P}^k。 \end{cases}$$

在对航班流量进行未来数据预测时，当前状态即历史数据的最后数值已知，记作 $\boldsymbol{\pi}(0)$。那么在该点上经过 k 个状态变化后，会有不同可能的状态概率，即 $\boldsymbol{\pi}(k)$。可以通过使用递推公式得到航班流量在第 k 个时间的状态。

同时在经过无穷多次状态转移后，所得到的状态概率可以预测在遥远的未来会出现的趋势。记概率向量为 $\boldsymbol{\pi}=[\pi_1,\pi_2,\cdots,\pi_n]$，则 $\pi_i=\lim_{k\to\infty}\pi_i(k)(i=1,2,\cdots,n)$，即有，$\lim_{k\to\infty}\pi_i(k)=\lim_{k\to\infty}\pi_i(k+1)=\pi$，将其带入马尔可夫预测模型的递推公式式(3)可得，$\boldsymbol{\pi}=\boldsymbol{\pi}\boldsymbol{P}$。其中状态概率应当满足的条件：

$$\begin{cases} \boldsymbol{\pi}=\boldsymbol{\pi}\boldsymbol{P}, \\ 0 \leqslant \pi_i \leqslant 1(i=1,2,\cdots,n), \\ \sum_{i=1}^{n}\pi_i=1。 \end{cases}$$

2 航班流量数据特点

2.1 航班流量的定义

机场航班流量通常指在某一个时间段内,从机场出发或者到达机场的航班数量。这个数量可以表示为进出港航班总数,也可以分别表示为起飞航班数量和降落航班数量。它不仅反映了航空运输业务的规模和发展趋势,还可以用来评估机场的设施和服务能力。

2.2 原始航班流量数据分析

本文以某机场实际运行的航班流量为例进行模型验证,以月为一个单位时间统计数据,将进场航班架次数与离场航班架次数之和作为航班流量,统计从 2018 年 5 月到 2022 年 2 月的航班流量,同时剔除了 2020 年 1 月到 6 月的数据,最终共 40 个月的数据,所得数据相对平稳,更加综合有效。若时间跨度范围太小,受不确定性因素的影响,会出现一些与平均值相差甚远的数值,影响整体的预测效果,从而失去了研究意义。为了便于表示,按照时间顺序给每个月标记序号,如图 1 所示。

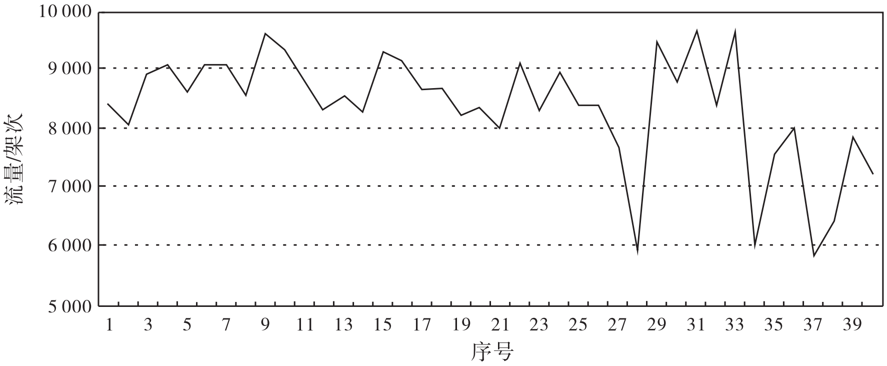

图 1 机场航班流量趋势图

通过观察趋势图可知,前 25 个月航班流量数据在一个稳定范围内上下浮动,且与时间之间呈现某种相关性,但由于受新冠肺炎中的各种随机扰动因素的影响,后续流量数据出现了下降和不平稳浮动的趋势。由于航班流量所受影响因素复杂且受到人们认识程度的限制,无法通过观察分析出准确的因果关系,所以建立合乎确定机理的数学模型,只能根据数据去建立模型,再根据数据去检验模型。

3 基于机场实际运行的航班流量预测

通过分析原始数据序列,发现航班流量呈现逐步下降的趋势,但受到随机扰动因素的影响出现了数据波动和偏离。因此本文分为两个部分进行预测分析:首先进行趋势性预测,对灰色预测模型和多项式回归的预测结果进行误差分析和比较,选择精度高的模型作为趋势性预测模型;其次使用马尔可夫模型对趋势性预测结果进行误差修正,比较修正前后的误差大小,并基于训练模型对测试集进行验证。

3.1 趋势性预测

3.1.1 *灰色预测* 针对灰色预测模型的适用条件,即小样本、变化规律不明显的特点,本文选取的 40 个月的航班流量数据相对于整个流量系统而言较少,可使用灰色模型进行预测。在对数据进行累加求和之后进行滑动平均处理,并根据实际数据建立灰色预测模型,其中模型所需参数可经计算得出,其中 $a=0.005\ 114\ 92$,$u=9\ 276.772\ 049$。同时计算历史数据预测值与真实值之间的相对误差,以及平均相对误差。将预测值与真实值数据进行比较,并计算残差,具体如图 2、图 3 所示。由图 2 可知,灰色预测模型可以大致反映出航班流量的整体趋势,其平均相对误差经计算为 1.19%。

3.1.2 *回归预测* 经计算,多项式回归曲线为 $y=8\ 570.877+50.221\ 89x-2.220\ 82x^2$,将拟合后的曲线与原始数据进行比较,计算残差,并在图中进行表示,具体结果如图 4、图 5 所示。同时计算拟合之后的

预测值与真实值之间的相对误差，以及平均相对误差。在数据拟合分析中观察到拟合曲线可以反映出原始数据的趋势，同时经计算所得，其平均相对误差为 1.07%。

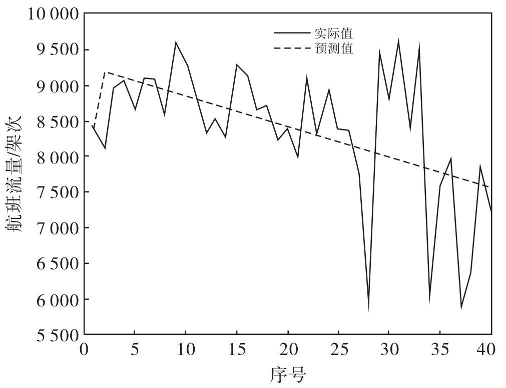

图 2　灰色预测模型实际值与预测值对比

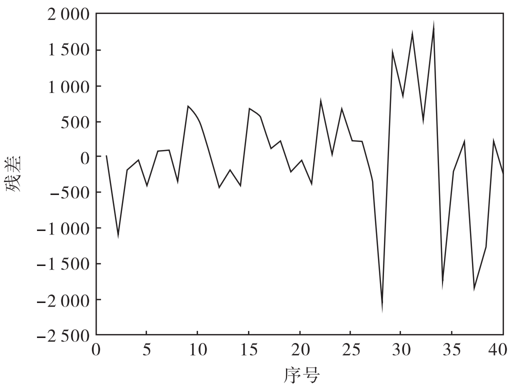

图 3　灰色预测模型残差图

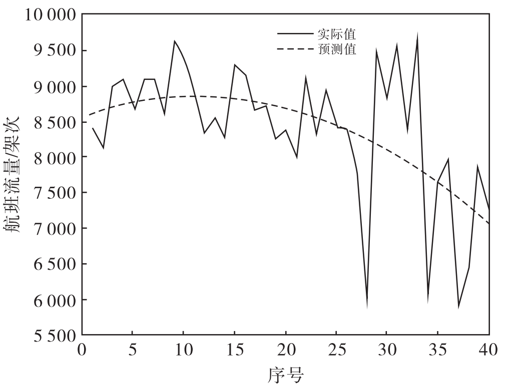

图 4　回归模型预测值与实际值对比

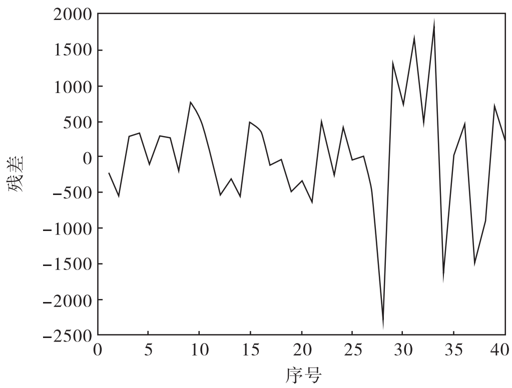

图 5　回归模型残差图

3.1.3　趋势模型选择　灰色模型和回归模型均能捕捉原始数据的趋势特征。比较发现，灰色预测模型的平均相对误差为1.19%，而回归模型的相对误差为 1.07%。鉴于此，本文选择回归模型作为主要的趋势性模型。但以上模型仅能揭示数据的大致趋势，而无法完全反映随机性质。为了更进一步地提高预测的准确性，需引入马尔可夫模型进行优化。

3.2　随机性预测

根据多项式回归模型的相对误差大小对航班流量进行状态分类，当相对误差小于－15%时为状态 1；当相对误差大小处于－15%～5%时为状态 2；当相对误差大于 5%时为状态 3。同时根据划分状态，对多项式回归拟合后的预测值进行误差修正，当处于状态 1 时，预测值上调 5%，当处于状态 3 时，预测值下调 5%，部分结果如表 1 所示。经计算可得，修正前平均相对误差为 1.07%，修正后为 0.64%，表明修正之后预测精度显著提高，如图 6 所示。

由图 6 可知，在回归模型反映出原始数据趋势的基础上，经马尔可夫模型修正后的数据可以反映出随机扰动因素的影响，且与原始数据重合度较高，表明了模型的可行性。

根据表 1 的状态划分，可计算出不同状态之间的转移概率，可得到航班流量的状态转移概率矩阵

$$\boldsymbol{P}=\begin{bmatrix}0.4545 & 0.3636 & 0.1818\\ 0.1111 & 0.5556 & 0.3333\\ 0.4 & 0.4 & 0.2\end{bmatrix}。$$

由表 1 可知，训练集中最后一个月的航班流量预测值与实际值之间的相对误差处于状态 2，则可根据状态转移概率矩阵，当前序数据处于状态 2 时，后序航班流量的相对误差大小应处于状态 2，表明相对误差的绝对值在 5%以内，无须修正。以 2022 年 3 月的航班流量作为测试集数据，其回归模型预测结果为

6879,实际航班流量为7156,相对误差为4.17%,处于状态2,验证了模型的可行性。

表1 马尔可夫修正过程

序号	回归预测结果与真实值相对误差	状态分类	马尔可夫修正结果与真实值相对误差	序号	回归预测结果与真实值相对误差	状态分类	马尔可夫修正结果与真实值相对误差
1	2.61%	2	2.61%	21	8.27%	3	2.85%
2	6.86%	3	1.52%	22	−5.51%	1	−0.78%
3	−2.95%	2	−2.95%	23	3.13%	2	3.13%
4	−3.71%	2	−3.71%	24	−4.98%	2	−4.98%
5	1.30%	2	1.30%	25	0.46%	2	0.46%
6	−3.26%	2	−3.26%	26	0.04%	2	0.04%
7	−2.86%	2	−2.86%	27	7.14%	3	1.79%
8	3.10%	2	3.10%	28	38.75%	3	31.81%
9	−8.11%	1	−3.51%	29	−14.04%	1	−9.74%
10	−5.11%	3	−9.86%	30	−8.05%	1	−3.45%
11	−0.16%	2	−0.16%	31	−17.35%	1	−13.22%
12	6.52%	3	1.19%	32	−5.46%	1	−0.73%
13	3.78%	2	3.78%	33	−19.14%	1	−15.09%
14	7.03%	3	1.68%	34	27.86%	3	21.47%
15	−5.01%	1	−0.26%	35	0.33%	2	0.33%
16	−3.62%	2	−3.62%	36	−5.94%	1	−1.23%
17	1.62%	2	1.62%	37	26.17%	3	19.87%
18	0.54%	2	0.54%	38	14.42%	3	8.70%
19	6.02%	3	0.72%	39	−9.07%	1	−4.52%
20	3.86%	2	3.86%	40	−2.67%	2	−2.67%

通过多项式回归模型的预测计算以及马尔可夫预测模型的进一步修正,模型可反映出原始数据的变化情况,且经过训练数据验证,相对误差在5%以内,表明模型的可行性。针对不同的空中交通流量情况,需要提前制定空中交通管制方案和地面保障程序,以此加快空中交通流动,提高管制效率,保障飞行安全。

4 结语

本文对机场航班流量分别进行趋势性预测和随机性预测,因机场航班流量在提前制定飞行计划的基础上还会受不确定性因素影响,且基于原始数据趋势图无法观察出其适配于哪一种常见曲线,因此选择灰色预测模型和多项式回归进行趋势性预测。比较二者预测后的平均相对误差,灰色模型为1.19%,回归模型为1.07%,因此选择回归模型预测数据作为趋势预测结果。然后使用马尔可夫模型对趋势预测结果进行误差修正,并根据状态转移验证测试集数据。结果表明,经修正后的模型平均相对误差为0.64%,且测试集数据误差在5%以内,验证了模型的可行性。故机场可根据此预测结果及时调整机场和空管的管制方案,进一步提高运行效率,保障运行安全。

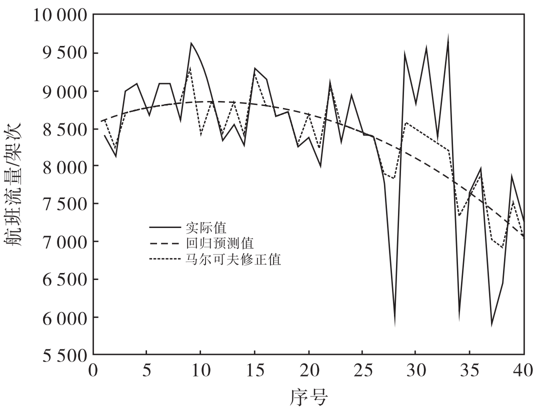

图6 实际值、回归预测值、马尔可夫修正值对比图

参考文献：

［1］ 向征，丁钰童．基于飞行计划集中处理的预战术航路飞行流量预测［J］．科学技术与工程，2021，21（24）：10525-10530．

［2］ 王兴隆，石宗北，贺敏．基于相似日 PSO-SVM 的机场流量预测［J］．计算机仿真，2022，39（7）：86-90．

［3］ 郭宪超．基于 ADS-B 数据的飞行流量预测［D］．广汉：中国民用航空飞行学院，2021．

［4］ 闫少华，谢晓璇，张兆宁．基于小波优化 GRU-ARMA 模型的空中交通流量短时预测方法［J］．交通信息与安全，2022，40（4）：177-184．

［5］ 尚然然．基于数据挖掘的终端区短期流量预测技术研究［D］．唐山：华北理工大学，2021．

［6］ KHAFIDLI M K，CHOIRUDDIN A．Forecast of Aviation Traffic in Indonesia Based on Google Trend and Macroeconomic Data using Long Short-Term Memory［C］//2022 International Conference on Data Science and Its Applications（ICoDSA）．IEEE，2022：220-225．

［7］ LU C Y，YANG L Y，LI S，et al．A Bayesian Network-Based Method for Air Traffic Flow Forecasting［C］//2023 IEEE International Conference on Sensors，Electronics and Computer Engineering（ICSECE）．IEEE，2023：1574-1577．

［8］ 朱承元，田睿．考虑管制员负荷的空中交通流量动态短期预测［J］．飞行力学，2024，42（2）：82-88．

［9］ 商博明，刘帅锋，王陕郡，等．基于灰色马尔可夫的市政道路路面使用性能预测［J］．公路，2022（3）：312-318．

［10］ 李剑，汪立新，白立伟．基于灰色马尔可夫预测的组合导航方法［J］．压电与声光，2023（1）：124-129．

［11］ 丁天明，潘宁，杜柏松，等．基于改进灰色马尔可夫的港口货物吞吐量预测研究［J］．重庆交通大学学报（自然科学版），2023，42（9）：130-136．

［12］ 王华友，张诚诚，薛海红，等．基于“回归＋马尔可夫”组合模型的民用航空事故症候数预测［J］．系统工程与电子技术，2023，45（7）：2114-2120．

Flight Traffic Prediction Based on Gray/Regression＋Markov Models

ZHAO Jiangge，ZHOU Jiemin

（*College of Civil Aviation*，*Nanjing University of Aeronautics and Astronautics*，*Nanjing* 211100，*China*）

Abstract：Reasonable and accurate flight flow forecast is the prerequisite for formulating relevant control plans and ground support procedures. A gray/regression＋Markov combination model is established to predict flight flow，and actual flight flow data at the airport are used as an example to verify the model. Firstly，the gray model and polynomial regression are used to predict the trend of the training samples of flight traffic，and the model with high accuracy is selected as the trend model. Secondly，the Markov model is used to correct the prediction results and the average relative error before and after the correction is compared. The results show that the error after correction is reduced from 1.07％ to 0.64％. Finally，the test sample is verified and the relative error is within 5％，which verifies the feasibility of the model. This research can provide a theoretical basis for timely adjustment of control plans，improvement of operational efficiency，and ensuring operational safety.

Keywords：flight traffic forecasting；gray forecasting；polynomial regression；Markov chains

（责任编辑：唐立平）

引用格式　赵江鸽，周洁敏．基于灰色/回归＋马尔可夫模型的航班流量预测［J］．山东航空学院学报，2024，41（1）：93-99．
ZHAO J G，ZHOU J M．Flight traffic prediction based on gray/regression＋Markov models［J］．Journal of Shandong University of Aeronautics，2024，41（1）：93-99．