

基于乘积季节模型的昆明市月度旅游人数预测

吕燕玲

云南财经大学, 云南 昆明
Email: 2451683602@qq.com

收稿日期: 2021年1月12日; 录用日期: 2021年2月18日; 发布日期: 2021年2月25日

摘 要

随着我国社会发展, 旅游产业扮演着越来越重要的角色。在有关旅游人数预测的研究中, 各种预测模型层出不穷, 本文借鉴现有的国内外研究经验, 以2013年1月~2020年10月昆明市月度旅游人数为依托, 运用时间序列模型中的乘积季节模型进行拟合, 对模型进行各项检验后利用该模型对昆明市未来两个月的旅游人数做了相应预测。

关键词

乘积季节模型, 旅游人数, 预测

Forecast of Monthly Tourist Numbers in Kunming Based on the Multiplicative Seasonal Model

Yanling Lv

Yunnan University of Finance and Economics, Kunming Yunnan
Email: 2451683602@qq.com

Received: Jan. 12th, 2021; accepted: Feb. 18th, 2021; published: Feb. 25th, 2021

Abstract

With the development of our society, tourism industry is playing an increasingly important role. There are all kinds of forecasting models in the studies of travel forecasting. Based on the existing research experience at home and abroad, this article uses the multiple seasonal ARIMA model to fit the monthly tourist population of Kunming from January 2013 to October 2020. Using this

model to predict the number of tourists in Kunming in the next two months when the model can pass all the tests.

Keywords

Multiplicative Seasonal Model, Number of Tourists, Forecast

Copyright © 2021 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. 研究背景及意义

昆明作为云南省省会，拥有独特的自然风景及得天独厚的气候条件，受到越来越多的人的青睐，也逐渐成为国内外游客旅游目的地第一选择。最近几年昆明市旅游业蓬勃发展，旅游业产值增长较快，据统计，在 2013 年昆明市接待国内游客共 5479.06 万人，实现旅游收入 490.92 亿元，在 2019 年接待国内游客 18,494.59 万人，全年旅游收入总计达到 2673.16 亿元。两项指标的年平均增长率分别为 22.48%、32.64%，每年旅游业的增加值在云南省 GDP 中的占比越来越大，逐渐发展成为云南的支柱产业之一。据测算，2019 年旅游产业综合增加值为 1101.85 亿元，旅游产业增加值对 GDP 的贡献达到 17.01%。同时，针对每一个省市地州旅游业的考核也已经纳入政府考核系统，因此各级政府都很重视旅游业的发展规划。要把旅游业发展好，除了推出新颖别致的旅游产品、加大对外宣传等传统手段外，更应该掌握市场上对当地旅游业的需求状况，从而加强本地区旅游业对市场需求变化的适应能力，了解本地旅游业的竞争趋势。这样才能做到有计划有目的的发展旅游业。

对旅游人数作出有效预测，有利于统筹安排当地的旅游资源及时调整景区浏览的承载力；有利于当地有关政府部门制定年度、月度旅游目标和计划；有利于调节旅游市场上方的供给方与需求方的平衡关系，合理安排好以及布局好当地旅游景点的开发战略以及游客的接待工作，使当地旅游部门发展更加高效率、高质量的旅游产业。

本文旨在探索出昆明市旅游人数发展变化的相关规律，构建和拟合出正确合理的实践系列模型，以期对昆明市未来旅游人数做出科学有效预测，也能及时了解旅游行业的规模及变化趋势，通过数据分析和预测能加强旅游相关规划目标的说服力。

2. 文献综述

在有关旅游市场方面的研究中，我国国内学者主要侧重于国内入境旅游需求方面的研究。按照研究方法可以大致分为定性研究方法、计量经济法(回归模型等)、时间序列法(ARIMA 模型、乘积季节模型、ARIMA-GARCH 模型等)、人工智能法(主要包含有神经网络、灰色理论方法等)。

在时间序列的相关预测方法中，陈鹏，吴玲，宋徽(2012) [1]曾基于时间序列方法中的 ARIMA 模型对安徽省入境旅游人数做了相应的预测，并经过分析比较得出了时间序列预测法在入境旅游人数预测中的应用具有良好效果的结论，同时指出短期内时间序列模型预测效果更好，所以若要进行长期预测则需要根据新的时间序列重新建立新的预测机制进行分析。孔朝莉(2018) [2]基于乘法季节模型对海南省旅游人数进行预测，在未来两期月度数据预测值中，过夜游游客预测值的平均绝对百分比误差为 1.69%，其中对于一期的预测得出的相对误差结果为 2.93%，第二期预测则为 3.39%。国内旅游人数预测值的平均绝对百分比误

差为 1.70%，其中对于一期的预测得出的相对误差结果为 2.88%，第二期预测则为 3.92%，说明应用本模型对于海南省未来两期的旅游人数预测是理想和合理的。同时也反映出季节模型适合短期预测，最好是一步预测。刘若愚，刘立波(2019) [3]基于 ARIMA 模型，对我国近几年的国内游客人数数据进行了模型拟合和预测，预测结果显示预测的误差在 1.4%左右，有良好的精度。虞安，王忠(2014) [4]基于引力模型及季节时间序列模型对第九届(北京)园博会参观人数进行预测，其思想为将游客按照来源地进行分类，然后分别采用旅游引力模型、季节性与非季节性自我回归整合移动平均模型以及本底趋势线模型进行预测。

在神经网络方法中，赵宇巍，张镇，郭子博，李远方基于 BP 神经网络的时间序列分析法对秦皇岛市旅游量进行预测研究，结果显示期望输出值与模型预测输出值之间的误差不足 0.1%，预测效果较好。而张锋，柳炳详，张月则运用 BP 神经网络算法建立时间序列模型对多个国家的旅游人数进行了拟合预测，结果表明该模型和算法能够对未来旅游人数进行有效预测。

此前邓聚龙(2005)提出了灰色系统理论和方法，这种方法主要适用于数量较少的数据。属于数列预测模型的一阶微分 GM(1, 1)模型作为现目前运用较多的灰色预测模型，王南丰，沈李欢(2020) [5]以陕西省 2011 年~2018 年旅游人数数据为基础，以 GM(1, 1)模型为基础对陕西省未来近五年的游客数量展开预测，根据最终预测结果，预测值相比实际值偏差较小，表明该模型能够科学合理地做出预测。程军锋也基于灰色理论用此模型对陇南市的旅游人数作了相应预测。郭况浦，李信韬(2019) [6]基于由灰色预测模型、线性回归模型组成的凸组合预测模型对 2008 年至 2017 年洛阳市旅游接待人数进行数据拟合，最终对洛阳市未来三年的国内旅游人数进行了预测分析。朱晓华、杨秀春和蔡运龙在灰色系统理论的基础上，建立合理的旅客来源灰色预测模型并对中国自 1978 年以来的入境客源数据进行拟合，同时比对了常用线性模型的预测结果，结果发现旅游系统适合运用灰色系统理论进行研究且有更高的预测精度。

在云南省旅游人数预测中，王政，杨霞(2020) [7]构建了灰色预测 GM(1, 1)模型，测算云南省各地州未来 6 年的入境旅游规模。除此之外有关云南入境旅游市场方面的预测研究尚难见到，而昆明市作为云南省的省会同时也是区域中心城市，对昆明市的旅游人数做出有效预测能够在昆明市旅游开发和旅游规划中起着至关重要的指导作用。上述文献综合了多种方法预测入境旅游人数，但并未得到一致结论，表明并不能找到适合所有实际情况的最优预测方法和模型。

通过上面的文献分析，以及拟使用的 2018 年 1 月~2020 年 10 月昆明市月度国内游客数据，根据选用的数据特征，本文将选用时间序列模型进行数据建模，但是具体采用什么模型则根据各项检验和估计来进行识别，选择所有能建立模型中的最优模型，最后通过建立的模型进行分析和预测。

3. 模型介绍

3.1. ARMA 模型介绍

当得到一个平稳的时间序列时，可以建立自回归移动平均模型即 ARMA(p, q)模型进行分析。若得到的序列是非平稳时间序列，则可对原序列进行差分提取确定性信息后再建立模型。此时可以对原序列建立 ARIMA(p, d, q)模型，模型中的 p 为自回归阶数， d 为消除非平稳序列需要差分的次数， q 为移动平均阶数。具体结构如下：

$$\begin{cases} \Phi(B)\nabla^d X_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \varepsilon_t^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_t \varepsilon_s) = 0, \forall s < t \end{cases} \quad (1) \text{式}$$

式中 $\nabla^d = (1-B)^d$, $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ ，为平稳可逆自回归移动平均模型的自回归系数多项式； $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ ，为自回归移动平均模型的平滑系数多项式。

3.2. 季节模型介绍

季节模型(seasonal model)是一类能够反映出具有季节变动规律的时间序列模型。时间序列的波动可以归结为四种因素的变动，分别是趋势因素、循环因素、季节因素随机变动因素。四种因素与原始序列的关系可用乘法模型和加法模型来表现。加法季节模型是指序列中的季节效应与其他效应是加法关系，即：

$$x_t = S_t + T_t + I_t \tag{2}式$$

简单季节模型实际上就是通过趋势差分、季节差分将序列转化为平稳序列，再对其进行拟合。它的模型结构通常如下：

$$\nabla_D \nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \tag{3}式$$

在时间序列的分析过程中，若简单的 ARIMA 模型并不足以提取各种效应之间的相关关系时，通常会采用乘积季节模型，比如季节变动的幅度随时间趋势的上升而增大这种情形，这时时序图的形状就会表现出前面阶段较小，后面阶段较大，即大致呈喇叭状或者放射状。

乘积模型的表达式如下：

$$\nabla^d \nabla_s^D x_t = \frac{\Theta(B)\Theta_s(B)}{\Phi(B)\Phi_s(B)} \varepsilon_t \tag{4}式$$

式中：

$$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\Phi_s(B) = 1 - \phi_1 B^s - \dots - \phi_p B^{ps}$$

$$\Theta_s(B) = 1 - \theta_1 B^s - \dots - \theta_q B^{qs}$$

该模型即为乘积季节模型，用符号简记为 ARIMA(p, d, q) \times (P, D, Q) $_s$ 。此时短期相关性和季节效应之间的关系为乘积关系，从表达式中也可看出。

3.3. 模型的检验

3.3.1. 模型的显著性检验

模型的检验可以分为对模型的显著性检验以及对参数的显著性检验。模型的显著性检验主要是用来判断建立的模型是否能有效的提取数据的有效信息，其检验的思想是若建立的模型能够有效，则模型中的残差序列应该为一个白噪声序列，此时残差中不存在任何有关原序列的信息，模型拟合的效果就会比较好。所以模型的显著性检验就转化为了残差序列的白噪声检，检验的原假设是残差为白噪声序列(模型显著)，备择假设是残差为非白噪声序列(模型不显著)。检验使用的统计量为 LB (Ljung-Box)检验统计量：

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{\rho_k^2}{n-k} \right) \sim \chi^2(m), \forall m > 0 \tag{5}式$$

同时也可以根据伴随概率 p 值判断模型是否显著。如果 LB 统计量的伴随概率 P 大于显著性水平，则认为这个拟合模型的残差序列属于白噪声序列，对应的模型显著有效。

3.3.2. 参数的显著性检验

参数显著性检验可以用来判断某一个参数显著还是不显著，即参数为零或者非零，同时通过参数的显著性检验可以是模型达到更加准确和精简的目的。该检验采用的统计量是 t 统计量并进行 t 检验，可通

过参数估计的 t 值与临界值相比较, 或者看该检验统计量的伴随概率来判断参数是否显著非零。具体地, P 小于显著性水平, 则对应的参数显著非零, 否则不显著。

3.3.3. 模型选择的依据

根据现有数据进行模型拟合时, 往往可以拟合出多个模型, 同时其中又有多个模型可以通过显著性检验。

4. 数据来源

本文选择的数据为 2013 年 1 月~2020 年 10 月的昆明市旅游人数月度数据, 数据来源于昆明市统计局官网发布的统计数据中的《各年月度昆明市国民经济主要指标》。

5. 实证分析

5.1. 原始数据平稳性检验

利用时间序列模型建模的基本要求是序列平稳且非白噪声序列。故首先判断本文选取的序列(将序列命名为 x)是否平稳, 得出的时序图及自相关图分别如图 1、图 2 所示。从时序图中可看出, 该序列 2013 年~2019 年期间都有明显的递增趋势, 2020 年由于受到疫情的影响, 出现了急剧下降。同时, 该时间序列还显示出明显的周期性特征。由自相关图 2 可看出该序列有严重的拖尾特征, 自相关系数一直都在零轴的上半部分, 没有截尾, 由此可判断该序列是不平稳的, 需要对原序列进行即一阶差分才能建模分析。

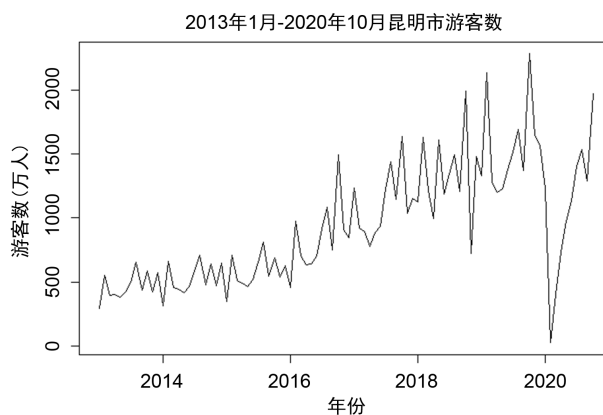


Figure 1. Sequence chart of series X

图 1. X 的时序图

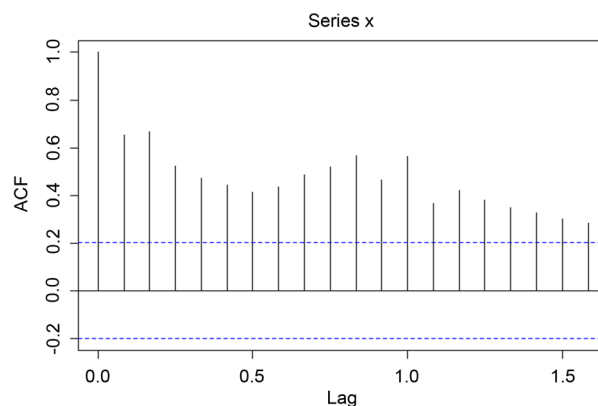


Figure 2. Autocorrelogram

图 2. 自相关图

5.2. 非平稳数据的处理及检验

由图 1 时序图中可看出昆明市游客数目前处于一个线性的增长趋势中，以及存在一定的周期性。再根据图 2 结果对原序列做一阶十二步差分后，由图 3 可知差分后的时间序列是平稳的。为了进一步精准判断时序数据是否平稳，对差分后的序列做单位根平稳性检验。检验结果如表 1 所示。检验结果显示统计量所对应的概率 P 小于 0.01，说明此时数据平稳。上文中指出，对于一个平稳的序列且为非纯随机的序列，才可以接着进行模拟和分析，否则就没有意义，所以根据时序图和检验结果应对差分后的序列进行纯随机性检验，得到滞后六期的 P 值为 0.0043，远小于 0.010 的显著性水平，此时数据为非纯随机，则可以进一步分析。

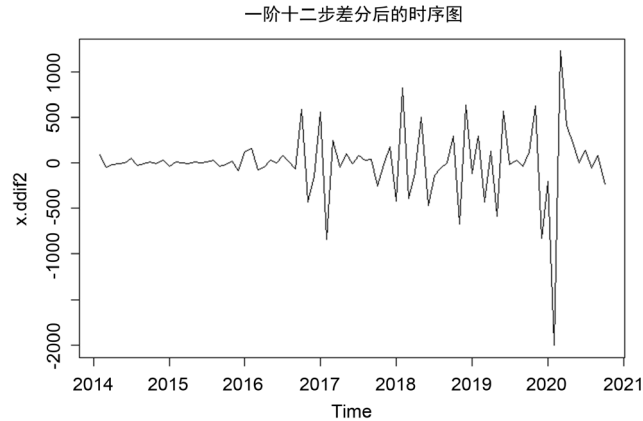


Figure 3. Sequence chart of difference series
图 3. 差分后的时序图

Table 1. Unit root stationarity test
表 1. 单位根平稳性检验

	DF 值	P 值
单位根平稳性检验	-5.8589	P < 0.01

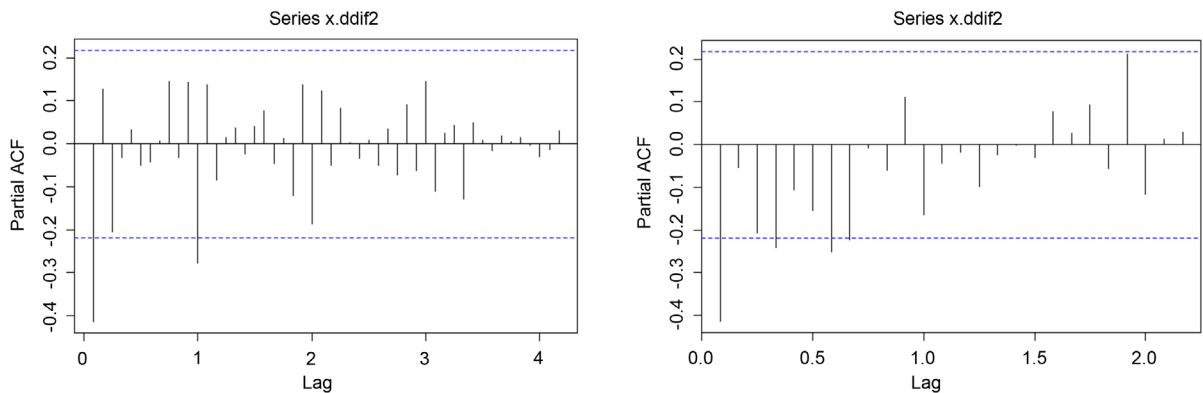


Figure 4. Autocorrelation and partial correlation
图 4. 自相关和偏自相关图

5.3. 模型的建立与检验

本文在应用模型对昆明市游客人数的预测过程中，根据图 4 中差分后序列的 ACF 图及 PACF 图分别拟合了如下模型：

Table 2. Comparison of models**表 2.** 模型的比较

模型	AIC 值	白噪声检验 P 值
模型一: ARIMA(1,1,0)×(1,0,1) ₁₂	1328.15	0.4341
模型二: ARIMA(1,1,1)×(0,1,1) ₁₂	1159.21	0.7198
模型三: ARIMA(4,1,1)×(0,1,1) ₁₂	1162.32	0.9944
模型四: ARIMA(7,1,1)×(0,1,1) ₁₂	1166.75	0.9979
模型五: ARIMA(8,1,1)×(0,1,1) ₁₂	1166.95	0.9999
模型六: ARIMA((1,4,7,8),1,1)×(0,1,1) ₁₂	1164.19	0.7993

根据表 2 中各模型的白噪声检验 P 值,发现上述模型都显著,根据 AIC 值大小比较发现模型二更优。模型显著后,进一步再检验了模型二残差的方差是否齐性,检验和判断模型二是否有 ARCH 效应,得到的结果如表 3 所示。

Table 3. ARCH test**表 3.** ARCH 检验

检验方法	滞后期数	P 值
LM 检验	1	0.682
	2	0.321
	3	0.5243
Q 检验	1	0.7843
	2	0.5996
	3	0.3607

由表 3 结果可知,模型二残差序列的 LM 和 Q 检验都认为方差是齐性的。

故 ARIMA(1,1,1)×(0,1,1)₁₂ 为本文根据数据特征建立的最优月度人数预测模型,其具体形式为:

$$\nabla \nabla^{12} x_t = \frac{1-0.8668B}{1-0.3543B} (1-0.6885B^{12}) \varepsilon_t$$

5.4. 模型的预测

此时 ARIMA(1,1,1)×(0,1,1)₁₂ 模型可以用来预测,在预测之前本文比较了该模型得出的 2020 年 9 月及 10 月的旅游人数预测值与真实值的预测误差百分比结果如表 4 所示。得到的结果显示模型的预测效果较好,接下来将用该模型预测昆明市 2020 年 11 月和 12 月的旅游人数,得到的结果如表 5 所示。

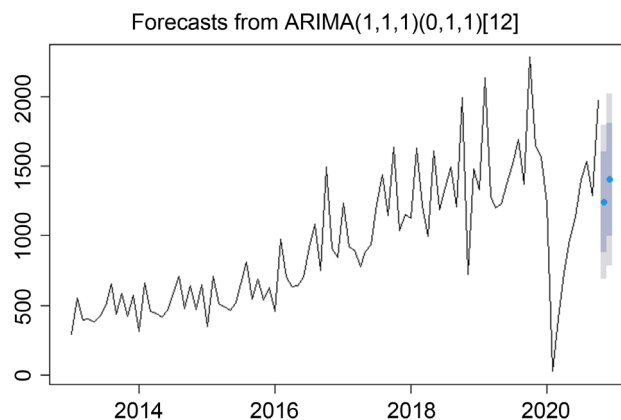
Table 4. Comparison between actual and predicted values**表 4.** 实际值与预测值的比较

	实际值	模型一预测值	误差百分比
2020 年 9 月旅游人数	1289.43	1209.907	6.79%
2020 年 10 月旅游人数	1974.07	1830.741	7.26%

Table 5. Prediction of the model**表 5.** 模型的预测

	2020 年 11 月游客数	2020 年 12 月游客数
预测值	1242.93 (万人)	1403.22 (万人)

预测图图 5 如下:

**Figure 5.** Prediction diagram**图 5.** 预测图

6. 结论及建议

本文基于 2013 年~月至 2020 年 10 月份昆明市旅游人数序列数据,建立了乘积季节模型 $ARIMA(1,1,1) \times (0,1,1)_{12}$ 对 2020 年 11 月及 12 月来昆旅游人数做了预测,预测结果为 2020 年 11 月份游客数为 1242.93 万人次,2020 年 12 月份游客数为 1403.22 万人次。相比刚过去的十月份,来昆旅游人数将有所下滑,不过,这符合历年数据的变化规律,十月份一般情况下都是第四季度接待游客量最多的月份,主要原因是十月份有一个国庆节这个长假,时间充足加上昆明市十月份气候较为温和,来昆旅游的人数自然会有大幅上升。到 12 月份后不少学校将会迎来寒假,此时来昆旅游人数相比 11 月份会有小幅上升。预测到 11、12 月份来昆旅游人数会有所下降,昆明市政府应该积极开展旅游推介会,加强昆明市旅游风光的宣传和展示,完善旅游设施配套提供更优质的旅游服务,吸引更多的国内外游客观光昆明市美景。

本文通过对旅游人数时间序列数据建立时间序列模型去模拟其变化情况,最终找到了一个旅游人数月度变化的具体形式,所以可以有效的预测未来旅游人数的变化趋势,有时候预测的具体数据可能不准确,但是预测到的变化趋势是符合现实情况的。从昆明市接待的游客的时序图中可以看出,2016 年之前昆明市每年的月度旅游人数变化趋势是一致的,说明每年的数据都是在上一年的基础上有一个固定的增幅,这不太符合社会发展的经济规律,所以在后面的相关预测中,容易受到这些数据的影响,预测的绝对数可能不是百分之百准确。但是大体的变化趋势是能预测到的。

参考文献

- [1] 陈鹏, 吴玲, 宋徽. 基于 ARIMA 模型的安徽省入境旅游人数预测[J]. 安徽农业大学学报(社会科学版), 2012, 21(1): 32-35+116.
- [2] 孔朝莉. 基于 ARIMA 模型的海南月度旅游人数预测[J]. 中国科技信息, 2019(5): 76-78.
- [3] 刘若愚, 刘立波. 基于 ARIMA 模型的游客人数分析与预测[J]. 电脑与电信, 2019(Z1): 1-4.

-
- [4] 虞安, 王忠. 基于 ARIMA 模型与时间序列的城市旅游倾向预测[J]. 统计与决策, 2014(13): 86-89.
- [5] 王南丰, 沈李欢. 基于 GM(1,1)模型的陕西省旅游人数预测[J]. 甘肃科技纵横, 2020, 49(5): 68-72+67.
- [6] 郭况浦, 李信韬. 基于综合预测模型的洛阳市旅游接待需求预测[J]. 内蒙古科技与经济, 2019(1): 27-28+33.
- [7] 王政, 杨霞. 云南省入境旅游规模时空分异特征及预测研究[J]. 云南农业大学学报(社会科学), 2020, 14(4): 93-102.