# Tourism Data Modeling and Mining based on Computer Information Technology

Yan-hong Sun[1a], Ting-ting Shang[2*]

[1]*Haust of Management, Henan University of Science and Technology, Luoyang, China*
[2]*Haust of Management, Henan University of Science and Technology, Luoyang, China*
[a]*zxgsyh126@126.com, *shangtt09@163.com*

**Abstract**
The wide application of computer information technology in the tourism industry has promoted the wave of tourism informatization, analyzed and excavated tourism data, explored new ways to integrate and optimize the allocation of tourism resources, and excavated wealth benefits. By building a spatial econometric model and combining the tourism data of 31 provinces and cities in China, this paper conducts an intelligent analysis of the spatial data and explores the spatial spillover effect. The results show that the tourism data of 31 provinces and cities in China has spatial relevance and the spatial spillover effect is significant.

***Keywords:*** *Computer Information Technology; Data mining and analysis; Spatial autocorrelation; Spatial econometric model*

## 1.INTRODUCTION

Nowadays, with the wide application of computer information technology, the Internet has become an indispensable part of people's life and also an important part of the development of the tourism industry. The interaction and integration between the Internet industry and the tourism industry lead to more diversified development modes. In the past, tourists lacked timeliness and accuracy in obtaining information, and the development of tourism was relatively slow. But now that the development of the Internet allows people to obtain information no longer has the limitation of distance and time, and the Internet has become the basis for the development of industrial integration and innovation.

By combing through the literature, foreign scholars have mainly analyzed the importance of tourism industry agglomeration and its role in promoting regional economic growth. Domestic scholars have mainly analyzed that tourism industry agglomeration has a significant role in promoting regional economic growth, and there is a high correlation and long-term co-integration relationship between the two. However, there are few researches on the agglomeration of the tourism industry from the perspective of the Internet. Zhu Xinglin analyzed the influence mechanism of the Internet on the regional tourism industry agglomeration from the perspective of supply and demand [10]. In recent years, with the improvement of the level of tourism industry agglomeration level and the growth of regional tourism economy, a few domestic scholars have studied the relationship between tourism industry agglomeration and tourism economic growth from the perspective of spatial spillover effects. Gao Jun et al. analyzed the impact of tourism industry agglomeration on tourism economy and its mechanism through theory and empirical analysis [2]. From the perspective of specialization and diversification, Wang Xinyue et al. found that the diversification of the tourism industry promotes the growth of the tourism economy, and the specialization agglomeration inhibits the growth of tourism economy [8]. Zhang Shuwen et al. analyzed the spatial spillover effect of tourism industry agglomeration on regional tourism economic growth based on the spatial Dubin model [11].

This study brings the Internet development level, tourism industry agglomeration and tourism economic growth into the research system, selects the data from 2004 to 2018, builds a spatial econometric model, and brings the core explanatory variables and control variables into the spatial lag term,a nalyzes and excavates tourism data, and explores the tourism spatial characteristics of 31 provinces in China.


G. Vilas Bhau et al. (Eds.): MSEA 2022, ACSR 101, pp. 794–799, 2023.
https://doi.org/10.2991/978-94-6463-042-8_113

## 2.SELECTION OF MEASUREMENT METHOD AND SETTING OF SPACE PANEL MODEL

### *2.1. Spatial Correlation Analysis*

Tobler argues that many things are related, and things that are closer are more related than things that are farther away. The same is true for economic data [7]. The closer the province is, the closer the connection is. Therefore, we should carry out the spatial correlation test to investigate whether the data has spatial dependence. This article is consistent with the practice of most scholars, using the Queen adjacency matrix binary method to construct spatial weight matrix. Due to the special geographical location of Hainan Province, it is assumed that it is only adjacent to Guangdong Province.

### *2.2. Spatial Econometric Model*

Panel data models are generally divided into fixed effects, random effects and mixed effects models, and the traditional spatial econometric model is mainly divided into SAR, SEM and SDM.

$$SAR: y = \alpha + \sum_j^n \rho W_{ij} y_{jt} + \beta X_{it} + c_i + \mu_t + \varepsilon_{it}$$

$$SEM: y = \alpha + \beta X_{it} + c_i + \mu_t + \sum_j^n \lambda W_{ij} v_{jt} + \varepsilon_{it}$$

$$SDM: y = \alpha + \sum_j^n \rho W_{ij} y_{jt} + \beta X_{it} + \sum_j^n \theta W_{ij} X_{jt} + c_i + \mu_t + \varepsilon_{it}$$

Among them, SEM adds a disturbance term to analyze the spatial effects of unobservable random shocks or missing variables. SAR is the most common in spatial econometric model, It is used to examine the spatial lag effect of dependent variables, that is, the influence of dependent variables in surrounding areas on local dependent variables. SDM is formed by the further development and evolution of SAR and SEM. At the same time, the spatial lag dependent variable and independent variable are included. It can measure the direct effects and spillover effects of dependent variables, and can effectively avoid the estimation bias caused by the omission of the spatial correlation of economic variables [4]. Due to space constraints, this article only lists the specific forms of SDM, as shown below.

$$\ln(pti)_{it} = \alpha + \sum_{j=1}^n \rho_1 W_{ij} \ln(pti)_{jt} + \beta_2 tia_{it} + \sum_{j=1}^n \rho_2 W_{ij} tia_{it} + \beta_3 ip_{it} + \sum_{j=1}^n \rho_3 W_{ij} ip_{it} + \beta_4 \ln(bli)_{it} + \sum_{j=1}^n \rho_4 W_{ij} \ln(bli)_{it} + \beta_5 \ln(irl)_{it} + \sum_{j=1}^n \rho_5 W_{ij} \ln(irl)_{it} + \sum_{j=1}^n \theta W_{ij} X_{jt} + c_i + \mu_t + \varepsilon_{it}$$

Where i represents the year, t represents the 31 provinces, W is the spatial weight matrix, ρ, α, β, θ are the coefficients to be estimated, $c_i$ represents the time fixed effect,, $\mu_t$ represents the individual fixed effect and $\varepsilon_{it}$ represents the random error term.

## 3.INDEX SELECTION AND DATA SOURCE

### *3.1. Index Selection*

On the basis of reviewing relevant literatures ([3][5][6][8][9][10][11]), this paper selects 9 indexs, which are explained variable, core explanatory variable and control variable (Table 1).

**Table 1:** Variable statistics table.

| Variable Category | Variable Name | Short Name |
|---|---|---|
| Explained variable | Tourism economy | pti |
| Core explanatory variables | Tourism industry agglomeration | tia |
| | Internet penetration rate | ip |
| | Internet basic level | bli |
| | Internet resource level | irl |
| Control variables | Economic growth | eg |
| | Industrial structure | str |
| | highway density | hd |
| | railway density | rd |
| | Regional openness | ro |
| | Regional innovation capability | ric |
| | Urbanization level | ul |

### *3.2. Data Source*

This paper selects the spatial panel data of 31 provinces in China from 2004 to 2018, and the data are from the "China Statistical Yearbook", "China City Statistical Yearbook", Statistical Yearbook of Provinces, National Bureau of Statistics, and "National Economic and Social Development Bulletin of Provinces". In order to eliminate the influence of price factors and ensure the comparability of data, the monetary indicators involved in this paper are all deflated and adjusted (using 2004 as the base period). In order to eliminate the influence of

heteroscedasticity, the non-percentage index is logarithmized in this paper.

## 4.EMPIRICAL RESEARCH

### *4.1. Spatial Correlation Test*

Based on the adjacency space weight matrix, Stata 15.1 software was used to calculate the global Moran index and local Moran index of tourism economy, tourism industry agglomeration and Internet development level. and plot the local Moran index scatter plot(Figure 1,Figure 2).

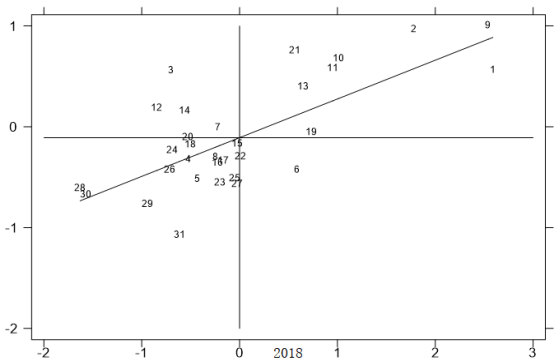


**Figure 1:** Part of the moran scatter plot of tourism economy.

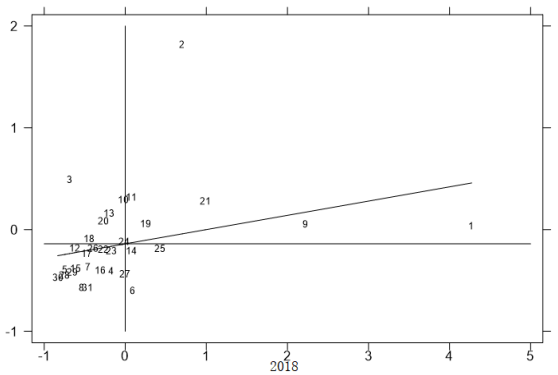


**Figure 2:** Part of the moran scatter plot of tourism industry agglomeration.

From 2004 to 2018, the global Moran index passed the 1%, 5%, and 10% significance level tests, indicating that there is a positive spatial correlation among the three, and the spatial clustering characteristics are obvious. There is also a significant positive spatial correlation between the local Moran indices, and the spatial locations are the "high-high" areas and the "low-low" areas.

### *4.2. Analysis of Spatial Panel Model Results*

#### *4.2.1. Model selection*

This paper refers to the detection method of Elhorst to select a spatial panel model with better fitting degree [1]. All statistics in LM and Robust LM tests have passed the significance level, which is suitable for the use of spatial econometric models. The results of the Hausman test (164.35, 0.0000) passed the 1% significance, rejecting the null hypothesis, so a fixed-effects model was selected. In the Wald test and the LR test, the P values of the two hypotheses were significant at the level of 0.01, and the goodness of fit of the SDM was significantly higher than that of the other two models, indicating that the SDM could not degenerate into SAR or SEM, so the SDM model was selected.

#### *4.2.2. Results analysis*

As shown in Table 2, it can be seen that the $R^2$ of the time-fixed-effect SDM model is more than 0.9, indicating its best fitting degree. The sign of the spatial lag term $\sigma^2$ is positive, and it has passed the significance test of 1%, which indicates that the tourism economy of 31 provinces in China shows a trend of agglomeration.

**Table 2:** Test results of spatial econometric model

| Variable | Time fixed | Space fixed | Time and space fixed |
|---|---|---|---|
| tia | 0.179*** | 0.251*** | 0.261*** |
| ip | 0.572*** | 0.102 | 0.134 |
| lnibl | -0.045*** | -0.001 | 0.000 |
| lnirl | 0.087*** | 0.015 | 0.013 |
| Wtia | -0.075* | -0.016 | 0.076 |
| Wip | 0.854** | -0.014 | 0.186 |
| Wlnibl | 0.002 | -0.007 | 0.000 |
| Wlnirl | -0.061 | -0.003 | -0.017 |
| $\sigma^2$ | 0.034*** | 0.003*** | 0.003*** |
| $R^2$ | 0.902 | 0.827 | 0.762 |

From the perspective of core explanatory variables, the correlation coefficient of tourism industry agglomeration level, Internet penetration rate and Internet resource level on tourism economy is positive, while the basic Internet level is negative. From the perspective of spatial lag variables, the tourism industry agglomeration and the Internet penetration rate respectively have a significant inhibitory effect and a promotion effect on the tourism economy of the surrounding areas with interactive connections in the region. The possible reason is due to the tourism industry agglomeration cannot form a synergistic effect through industrial associations between regions, but the Internet allows tourism information without regional restrictions, which greatly stimulates the tourism motivation in the surrounding areas and promotes the generation of tourism activities. Wlnibl, Wlnirl have no significant effects, indicating that their impact on the tourism economy in surrounding areas has not been manifested.

### *4.3. Spillover Effect Analysis of Spatial Panel Model*

In view of the above model estimation results, the spatial Dupin time-fixed effect model is selected to decompose the spatial spillover effect using the partial differential method of the spatial regression model (Table 3).

(1) Direct effect.

The spatial correlation coefficient of tourism industry agglomeration, Internet penetration rate, Internet resource level, economic growth, industrial structure, highway density, and urbanization is positive, and has passed the 1% significance test, indicating that it has a promoting effect on the tourism economy. The direct effect of industrial structure on the tourism economy is greater than the direct effects of other variables. As for traffic factors, only the impact of railway density on the development of tourism economy is inhibited. The main reason may be that the increase in railway traffic speed attracts tourists to go to farther places. The spatial correlation coefficient of regional openness and regional innovation ability did not pass the significance test.

**Table 3:** Spatial spillover effect decomposition

| Variables | LR_Direct | LR_Indirect | LR_Total |
|---|---|---|---|
| tia | 0.185*** | -0.099*** | 0.086** |
| ip | 0.532*** | 0.623* | 1.155*** |
| lnibl | -0.044*** | 0.010 | -0.034 |
| lnirl | 0.090*** | -0.072 | 0.018 |
| lneg | 0.191*** | 0.188*** | 0.379*** |
| str | 2.178*** | 4.448*** | 6.626*** |
| lnhd | 0.268*** | -0.250*** | 0.019 |
| lnrd | -0.271*** | 0.165*** | -0.106* |
| ro | -0.021 | 0.571*** | 0.550*** |
| lnric | 0.020 | -0.162*** | -0.142*** |
| ul | 2.176*** | -0.853* | 1.323** |

(2) Indirect effect. Except that the spatial correlation coefficient between the basic level of the Internet and the level of Internet resources did not pass the significance test, the rest of the variables passed the 1% and 5% significance tests. The tourism industry agglomeration has a significant negative spatial spillover effect on the tourism economy of other regions, that is, it has an inhibitory effect on the growth of the tourism economy in the surrounding areas. The indirect effect of the industrial structure is much higher than the direct effect, indicating that in the context of comprehensive tourism, the optimization of the industrial structure can produce diffusion effects and drive the growth of tourism economy in surrounding areas. The spatial correlation coefficient between highway density and tourism economy is negative, which may be due to the "siphon effect". The enhancement of local highway infrastructure has attracted tourists from surrounding areas, thus hindering the growth of tourism economy in surrounding areas. Regional innovation capacity and urbanization have an inhibitory effect on the growth of tourism economy in surrounding areas, the main reason may be that the increase in the level of urbanization in surrounding areas makes it competitive with the local area in human capital and infrastructure.

(3) Total effect.The total effect of Internet basic level, Internet resource level and highway density is not significant, and the other variables have passed the significance test of different degrees. The correlation coefficient of the total effect of tourism industry agglomeration is positive. Due to the total effect is equal to the sum of direct effect and indirect effect, the direct effect of the tourism industry agglomeration on the tourism economy is higher than the indirect effect, indicating that although the tourism industry agglomeration can significantly promote the growth of the tourism economy in the region, it has a smaller negative spillover effects on the surrounding areas. On the whole, the tourism industry agglomeration is beneficial to tourism economic growth, and regional tourism activities have not formed a good mutual promotion and mutual assistance effect. The correlation coefficient of the total effect of the Internet penetration rate is positive, indicating that the development of the Internet promotes the expansion of the tourism industry, promotes the development of smart tourism and comprehensive tourism, and expands the tourism consumption circle.

## 5.CONCLUSIONS AND SUGGESTIONS

### *5.1. Conclusions*

(1) There is a significant positive spatial correlation in the tourism economy, tourism industry agglomeration, and Internet development level in various provinces. Most regions are located in the "high-high" area and "low-low" area. Therefore, the spatial effect cannot be ignored when exploring the tourism industry agglomeration and tourism economy.

(2) The tourism economy has a significant spatial spillover effect. The agglomeration of tourism industry and the density of highways have a positive effect on the tourism economy of the region, but they have a negative effect on the surrounding areas. The density of railways is just the opposite. Internet penetration rate, economic growth, and industrial structure have all contributed to the tourism economy in this region and surrounding areas. The basic level of the Internet has a negative inhibitory effect on the tourism economy of the region, and the level of Internet resources and urbanization has a positive inhibitory effect on the tourism economy of the region, but has no significant effect on the surrounding areas. The degree of regional openness has no significant effect on the tourism economy of the region, but has a positive effect on the surrounding areas.

(3)The impact of related variables on the tourism economy has direct and indirect effects. The indirect effects of the Internet penetration rate, industrial structure, regional openness and regional innovation ability are greater than the direct effects, and the other variables are that the direct effect are greater than the indirect effects, indicating that the tourism economy is simultaneously affected by the development of the region and the surrounding areas, but the region is the main influence factor. From the perspective of total effect, in addition to the Internet basic level, the tourism industry agglomeration, Internet penetration rate, economic growth, industrial structure, regional opening level and urbanization level have positive promoting effects on the tourism economy.

### *5.2. Suggestions*

(1) Due to the positive spatial correlation and spatial spillover between the tourism industry agglomeration, the level of Internet development and the tourism economy, the coordination of tourism industries among various regions should be brought into play to create tourism belts and tourism central cities, driving low level areas through high level areas, improve the level of tourism industry agglomeration and then promote tourism economy growth. At the same time, all regions should take advantage of the Internet, such as big data, connection and sharing, etc., to promote the inter-regional "Internet + tourism" model, realize the joint development and shared development of all regions, and form a spatial synergy, so as to promote the growth of tourism economy. All regions should make use of cloud computing, big data and other technologies to build tourism industry clusters. On the basis of continuous improvement of tourism public information services, mobile terminals such as smart phones should be made full use of to meet the needs of tourists. The tourism industry and enterprises should actively integrate information technology, connect the upstream and downstream industrial chains and value chains, and promote the coordinated development of the tourism industry and enterprises. Tourism management departments and relevant government departments should optimize the institutional environment, establish the correct value concept and system, promote the agglomeration of tourism industry, and promote the growth of tourism economy.

(2) Although the tourism industry agglomeration can promote the growth of the local tourism economy, it has inhibiting effect on the tourism economy of the surrounding areas. In the process of promoting tourism industry agglomeration, each region should have a clear understanding of its own tourism resource advantages, scale characteristics and development stage, formulated, adjusted and implemented according to local conditions. On the basis of the implementation of the development concept of "comprehensive tourism", promote the coordinated configuration of "tourism+" and "+tourism",continue to increase the degree of integration between the tourism industry and related industries, and optimize the industrial structure. Local governments should pay attention to the construction of tourism industry agglomeration area, but should avoid excessive competition in the same tourism industry agglomeration area.

## ACKNOWLEDGMENTS

The Major Peoject of Basic Research on Philosophy and Social Sciences in colleges and unicersities in Henan Province,2021 “Research on the perception of all-for-one tourism image in Henan Province”.

The successful completion of the dissertation is inseparable from the encouragement and support given to me by Professor Sun and Senior Sister Wang Qingnan. I am very grateful to them.

## REFERENCES

[1] ELHORST J P.Matlab Software for Spatial Panels[J].International Regional Science Review,2014,37(3):389-405.

[2] Gao J, Zhang L L. Study on relationship of China's tourism industry agglomeration, total factor productivity and tourism economy[J]. Resource Development & Market, 2017, 33(8):1005-1010.

[3] Huang B M. Whether economic agglomeration can kill two birds with one stone: Spatial econometric analysis of growth effect and energy conservation and emission reduction effect[J].Modern Economic Research,2020(07):20-32.

[4] LEE L F,YU J.Some recent developments in spatial panel data models[J].Regional Science and Urban Economics,2010,40(5):255-271.

[5] Liu J, Zhao J J, Zhang G H. Spatial econometric research on the relationship between tourism industry agglomeration and tourism economic growth in China[J]. Economic Geography,2013, 33(4):186-192.

[6] Shao S, Zhang K, Dou J M. Effects of economic agglomeration on energy saving and emission reduction: Theory and empirical evidence from China[J]. Management World, 2019,35(01):36-60+226.

[7] Tobler W R. A Computer Movie Simulating Urban Growth in the Detroit Region[J]. Economic Geography,1970,46.

[8] Wang X Y, Lu X J. Evolution of China’s tourism industry agglomeration spatial pattern and its

impact on tourism economy based on specialization and diversification agglomeration perspective[J]. Scientia Geographica Sinica, 2020,40(7):1160-1170.

[9] Wang M K, Liu Y P, Li T. The differential impact of tourism industrial agglomeration on environmental pollution: Empirical evidence from 287 cities. Reform, 2019(2):102-114.

[10] Zhu X L.An empirical study on the influence of Internet on tourism industry agglomeration in China[D].East China Normal University,2020.

[11] Zhang S W, Chen Q C, Wang K. Relationship between tourism industry agglomeration and regional tourism economic growth: Based on China's Provincial Panel Data during 2001-2017[J]. Tropical Geography,40 (1):154-163.