# Spatio-temporal analysis of influenza-like illness and prediction of incidence in high-risk regions, in the United States from 2011 to 2020

**Zhijuan Song**
Zhengzhou University
**Xiaocan Jia**
Zhengzhou University
**Junzhe Bao**
Zhengzhou University
**Yongli Yang**
Zhengzhou University
**Huili Zhu**
Zhengzhou University
**Xuezhong Shi** ( xzshi@zzu.edu.cn )
Zhengzhou University

**Research Article**

**Keywords:** influenza-like illness, spatio-temporal analysis, SARIMA model, prediction

**Posted Date:** February 23rd, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-220805/v1

## Abstract

# Introduction:

About 8% of Americans get influenza during an average season from the Centers for Disease Control and Prevention in the United States. It is necessary to strengthen the early warning of influenza and the prediction of public health.

# Methods

In this study, we analyzed the characteristics of Influenza-like Illness (ILI) by Geographic Information System and SARIMA model, respectively. Spatio-temporal cluster analysis detected 23 clusters of ILI during the study period.

# Results

The highest incidence of ILI was mainly concentrated in the states of Louisiana, District of Columbia and Virginia. The Local spatial autocorrelation analysis revealed the High-High cluster was mainly located in Louisiana and Mississippi. This means that if the influenza incidence is high in Louisiana and Mississippi, the neighboring states will also have higher influenza incidence rates. The regression model SARIMA(1, 0, 0)(1, 1, $0)_{52}$ with statistical significance was obtained to forecast the ILI incidence of Mississippi.

# Conclusions

The study showed, the ILI incidence will begin to increase in the 45th week 2020 and peak in the 6th week 2021. To conclude, notable epidemiological differences were observed across states, indicating that some states should pay more attention to prevent and control respiratory infectious diseases.

## Introduction

Influenza is caused by influenza virus which mainly spreads through airborne droplets and direct contact. It has the characteristics of strong infectivity, rapid transmission virus and antigen variation. The activity of influenza begins to increase in October, most often peaks between December and February and can remain elevated into May. Influenza virus infections are very common and their incidence can only be estimated.[1] About 8% of Americans get influenza during an average season from the Centers for Disease Control and Prevention (CDC).[2] In particular, the 2017–18 influenza season in the United States was known for its high severity, with about 45 million illnesses and 810,000 influenza-associated hospitalizations.[3] Therefore, it is necessary to strengthen the early warning and public health forecasting of influenza.[4]

Since influenza might be with fever, cough, sore throat, runny or stuffy nose, body aches, headache, chills or fatigue and so on, it's hard to be diagnosed as influenza based on symptoms alone. A number of influenza tests are available to detect influenza viruses in respiratory specimens. The most common are called "rapid influenza

diagnostic tests (RIDTs)” .[5] The most important thing is that not all people got tested for influenza. So, the number of reported cases of influenza may significantly underestimate the actual prevalence of influenza. In order to more accurately understand the current situation of influenza, this study conducted a spatio-temporal analysis by Influenza-like Illness (ILI).

ILI transmission occurs locally and is spread from person to person,[6] and the prevalence of influenza varied among regions in the United States. For example, the Southeastern United States influenza estimates were higher than the Northwestern United States estimates of influenza. The Geographic Information System (GIS) is a useful epidemiological method for identifying high-risk areas in many infectious diseases.[7] Identifying high-risk and endemic areas and the spatial-temporal distribution of infectious diseases is important for the development of prevention plans and health policies.[8]

In the temperate regions, ILI is predominantly seasonal.[9] For diseases that show recurrent seasonal patterns or occur in cyclic patterns, time series models are the most widely used statistical models by health researchers for forecasting.[10] The autoregressive integrated moving average (ARIMA) is a widely used predictive analysis model for non-stationary time series,[4] which is of great significance to the study of non-stationary time series. In this study, the model was used to predict ILI outbreaks, and combined with spatial geographic analysis to provide theoretical guidance and scientific basis for the prevention and treatment of influenza.[11]

In the absence of similar research on ILI, the aim of this study was to identify the hotspots states in the United States and to predict the incidence of ILI in hotspots, which can help health policymakers to better plan for disease prevention and control in high-risk states.

## Materials And Methods

# Data Resources

Information on outpatient visits to health care providers for ILI is collected weekly through the United States Outpatient Influenza-like Illness Surveillance Network. For this system, the confirmed influenza case was “A patient who tests positive for influenza virus infection by an approved laboratory test”, and ILI is defined as “fever (temperature of 100°F (37.8°C) or greater) and a cough and/or a sore throat without a known cause other than influenza” (https://www.cdc.gov/flu/weekly/overview.htm).

We collected the CDC’s weekly ILI data from 1st week 2011 to 29th week 2020. The data included the number of ILI cases in 49 states for different age groups. For Hotspot States, time series models were constructed by collecting data from 1st week 2011 to 52nd week 2018, and data from 1st week 2019 to 29th week 2020 were taken as test data to assess forecast performance.[2]

# Spatio-Temporal Cluster Analysis

Spatio-Temporal Cluster Analysis is a measurement of temporal and spatial correlation on the foundation of spatial autocorrelation with the further consideration of the time factor.[12] It can relate spatial characteristics to temporal characteristics of influenza.[11] During the study period, the cluster was detected by retrospective spatio-temporal permutation scanning statistics.[13] After detecting the most likely spatio-temporal clusters, these clusters were tested by the Monte Carlo method. If the points conforming to the evaluated cluster

maintained their aggregated pattern when compared with 999 randomized simulations of the entire dataset, then it was considered important.[14]

Moran's I is an important index for analyzing the spatial distribution of diseases, and has been used in many studies. The value of Moran's I ranges from-1 to 1, where 0 indicates a random distribution of influenza. A value close to 1 indicates that the unit cluster has a similar value. A value close to -1 indicates that the unit with high values and low values are adjacent in space.[15] Based on its value and significance, Moran's I can detect four types of clustering, reflecting the high-high (HH), high-low (HL), low-low (LL) and low-high (LH) clustering patterns, respectively. The number of permutation was 999, and the significance level was 0.05.[16]

Unlike spatial autocorrelation, scan statistics can evaluate the aggregation of observations and the location of aggregated observations. Spatio-temporal scan statistics are defined by a specific window with a circular geographic base and height corresponding to time. The window size was constantly adjusted to detect possible spatio-temporal clusters.[13] In order to scan for small to large clusters, the largest radius was set to 50% of the total population at risk, the largest height was set to 50% of the total study period.[17]The Logarithmic Likelihood Ratio (LLR) was used to compare observed and expected numbers to identify specific clusters. The most likely clusters were determined by the Monte Carlo test.[18]

# Time-Series Analysis

Time series analysis has the advantage of predicting incidence. It is characterized by the number of patients in the past and responds by predicting the number of patients in the future. The ARIMA model is based on the sequential lag relationship existing in time-series data.[4] However, the SARIMA model is more suitable for forecasting when the data has obvious seasonal characteristics. The SARIMA model can be expressed as: SARIMA(p, d, q)$(P, D, Q)_s$. Letters p, d, q is the order of autoregression, the order of difference and the order of moving average; Letters P, D, Q are the order of seasonal autoregression, the order of difference and the order of moving average, and s is the specific value of cycle, the cycle of American influenza is 52 weeks(s = 52).[4]

The process of establishing the SARIMA model is divided into three steps: First, a weekly time-series plot of incidence(per 100,000 populations) was drawn to check for stationarity and seasonality. The model was constructed according to the auto-correlation function (ACF) and partial auto-correlation function (PACF) of the model residuals. Secondly, ACF and PACF for estimating residuals are tested by Ljung-Box Q test, and the minimum of the Bayesian information criterion (BIC) is taken as the optimal SARIMA model. Finally, the model is applied to forecast the weekly incidence for 30th week 2020 to 52nd week 2021.

# Statistical Analysis

The data was organized by Microsoft Excel 2013. The SARIMA model was created by Rstudio and SPSS27.0. The value of Moran'I and local indicators of spatial association were calculated by GeoDa1.14.0. The time scan statistic was measured with SaTScan$^{TM}$9.5. All the maps were drawn by ArcGIS 10.0.

## Results

# Epidemiological analysis

Included in our study were a total of 9,065,910 ILI cases from 1st week 2011 to 29th week 2020 among the United States. The ILI annual infection rate fluctuated from 5.92 to 15.84 per 100,000 populations. ILI occurs throughout the year, most often peaking between December and February and can last until May.

In terms of age, the number of ILI cases in the age 5–24 years old is the most, and these groups accounted for about 35 percent, while the number of patients in the age group over 65 years old is the least, it accounted for about 7 percent (Fig. 1). The difference between different age groups had a statistical significance ($P < 0.001$).

This study collected the population of 49 states and visualized them on the map (Fig. 2). Comparing these results with spatio-temporal analysis could reveal an association between population density and influenza incidence.

# Spatio-Temporal Analysis

Overall, the highest cumulative incidence of ILI incidence (per 100,000 populations) during the study period was seen in the states of Louisiana, District of Columbia and Virginia of reported cases 12,200, 9,563 and 9,554, respectively. The lowest cumulative incidence of ILI incidence was reported from the states of Ohio, Washington and Iowa (Fig. 3).

# Global Spatial Autocorrelation

The global spatial autocorrelation analysis for ILI suggested a clustering distribution at the state level in the years of 2012 to 2017, the global Moran's I reached up to the significance level of 0.05. In contrast, the global Moran's I for 2011, 2018 and 2019 display no significant spatial autocorrelation, though Moran's I greater than 0 (Table 1).

Table 1
Global spatial autocorrelation analysis

| Year | Moran's I | E(I) | Mean | s | Z-Value | P-Value |
|---|---|---|---|---|---|---|
| 2011 | 0.099 | -0.021 | -0.023 | 0.090 | 1.358 | 0.093 |
| 2012 | 0.185 | -0.021 | -0.019 | 0.095 | 2.151 | 0.028 |
| 2013 | 0.181 | -0.021 | -0.019 | 0.094 | 2.121 | 0.031 |
| 2014 | 0.200 | -0.021 | -0.021 | 0.092 | 2.401 | 0.022 |
| 2015 | 0.166 | -0.021 | -0.222 | 0.091 | 2.073 | 0.037 |
| 2016 | 0.177 | -0.021 | -0.024 | 0.093 | 2.150 | 0.029 |
| 2017 | 0.146 | -0.021 | -0.025 | 0.087 | 1.981 | 0.039 |
| 2018 | 0.074 | -0.021 | -0.023 | 0.092 | 1.053 | 0.145 |
| 2019 | 0.074 | -0.021 | -0.021 | 0.090 | 1.053 | 0.139 |

# Local Spatial Autocorrelation

Local spatial autocorrelation analysis reveals only the relative states, rather than absolute correlations. Only those states whose local Moran's I have reached the significance level of 0.05 will present on the map. From 2011 to 2019, the local spatial autocorrelation showed 3 HH clusters in total with 2 HL clusters, 4 LH and 3 LL clusters. HH clusters were observed in the states of Louisiana (5 years), Mississippi (4 years), and the District of Columbia (1 year). Louisiana and Mississippi had HH clusters for long periods. HL clusters were observed in the states of Illinois (4 years), and Oregon (2 years). LH clusters were observed in the states of Tennessee (4 years), Maryland (6 years), Arkansas (5 years), and Texas (3 years). LL cluster appears in the northeastern part of the United States only in 2018 and 2019 (Fig. 4).

# Spatio-Temporal Cluster Analysis

The spatio-temporal Cluster analysis detected 23 clusters of ILI in the study period. The clusters were particularly obvious in the spring and winter. For example, Risk Ratio (RR) was highest in 2015, with a total of 3 levels of clustering. Level 1, with Louisiana at the center of high incidence area and 2 surrounding states, the risk of ILI in this area was 11.66 times more likely to develop the disease than other areas ( LLR = 69,009, $P <$ 0.001). Level 2, with Virginia at the center of high incidence area and 3 surrounding states, the risk of ILI in this area was 9.79 times more likely to develop the disease than other areas ( LLR = 73,277, $P <$ 0.001). Level 3, with New Mexico at the center of high incidence area and 3 surrounding states, the risk of ILI in this area was 3.38 times more likely to develop the disease than other areas ( LLR = 26,518, $P <$ 0.001). At the same time, the states with high cluster in the Local spatial autocorrelation analysis were all located in the high cluster area, the results were consistent. From the cluster time, the high incidence time mainly occurs between January and March (Table 2).

**Table 2.** Spatio-temporal scan of ILI in the United States from 2011 to 2019.

| Year | Level | Center | N | Cluster period | Coordinates/Radius(km) | Observed cases | Expected cases | RR | LLR | P-value |
|---|---|---|---|---|---|---|---|---|---|---|
| 2011 | 1 | Kentucky | 15 | 2011-01-01 to 2011-02-28 | (37.5N,85.3W)/738.2 | 121829 | 24485 | 6.22 | 108601 | <0.001 |
| | 2 | Colorado | 13 | 2011-01-01 to 2011-02-28 | (39.0N,105.5W)/1005.2 | 62527 | 15984 | 3.91 | 4.32 | <0.001 |
| 2012 | 1 | Mississippi | 3 | 2012-10-01 to 2012-12-31 | (32.8N,89.7W)/289.2 | 37831 | 4698 | 8.65 | 46953 | <0.001 |
| | 2 | Virginia | 1 | 2012-09-01 to 2012-12-31 | (37.5N,78.8W)/0 | 29130 | 4234 | 7.26 | 31944 | <0.001 |
| | 3 | Nebraska | 17 | 2012-01-01 to 2012-03-31 | (41.5N,99.8W)/1116.02 | 65301 | 32693 | 2.15 | 13777 | <0.001 |
| 2013 | 1 | Virginia | 3 | 2013-01-01 to 2013-03-31 | (37.5N,78.8W)/222.1 | 46638 | 4750 | 10.61 | 66247 | <0.001 |
| | 2 | Texas | 15 | 2013-01-01 to 2013-2-28 | (31.5N,99.4W)/1327.5 | 101773 | 27333 | 4.32 | 64734 | <0.001 |
| 2014 | 1 | Mississippi | 3 | 2014-10-01 to 2014-12-31 | (32.8N,89.7W)/289.2 | 45125 | 5686 | 8.52 | 55411 | <0.001 |
| | 2 | Virginia | 3 | 2014-01-01 to 2014-04-30 | (37.5N,78.8W)/222.1 | 43127 | 6524 | 7.06 | 46035 | <0.001 |
| | 3 | New Mexico | 12 | 2014-01-01 to 2014-02-28 | (34.4N,106.1W)/1249.5 | 54639 | 18361 | 3.18 | 24494 | <0.001 |
| 2015 | 1 | Louisiana | 2 | 2015-01-01 to 2015-04-30 | (31.1N,92.0W)/289.2 | 45837 | 4254 | 11.66 | 69009 | <0.001 |
| | 2 | Virginia | 3 | 2015-01-01 to 2015-04-30 | (37.5N,78.8W)/222.1 | 54666 | 6134 | 9.79 | 73277 | <0.001 |
| | 3 | New Mexico | 12 | 2015-01-01 to 2015-02-28 | (34.4N,106.1W)/1249.5 | 54311 | 17264 | 3.38 | 26518 | <0.001 |
| 2016 | 1 | Virginia | 3 | 2016-01-01 to 2016-05-31 | (37.5N,78.8W)/222.1 | 63287 | 7974 | 8.81 | 78624 | <0.001 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | Arizona | 3 | 2016-01-01 to 2016-04-30 | (34.3N,111.7W)/559.1 | 32319 | 7155 | 4.73 | 24144 | <0.001 |
| | 3 | Mississippi | 10 | 2016-01-01 to 2016-04-30 | (32.8N,89.7W)/820.6 | 104314 | 34608 | 3.47 | 50171 | <0.001 |
| 2017 | 1 | Florida | 13 | 2017-01-01 to 2016-03-31 | (28.6N,82.5W)/1247.2 | 192625 | 51339 | 4.63 | 127775 | <0.001 |
| | 2 | Oregon | 1 | 2017-11-01 to 2017-12-31 | (43.9N,120.6W)/0 | 6819 | 1714 | 4.01 | 4329 | <0.001 |
| | 3 | Colorado | 14 | 2017-01-01 to 2017-02-28 | (39.0N,105.5W)/1024.3 | 54139 | 25185 | 2.23 | 13031 | <0.001 |
| 2018 | 1 | Florida | 13 | 2018-01-01 to 2018-02-28 | (28.6N,82.5W)/1247.2 | 259855 | 47137 | 6.89 | 253645 | <0.001 |
| | 2 | Wyoming | 13 | 2018-01-01 to 2018-02-28 | (43.0N,107.6W)/1052.7 | 74812 | 19512 | 4.04 | 46665 | <0.001 |
| 2019 | 1 | Colorado | 7 | 2019-01-01 to 2019-03-31 | (39.0N,105.5W)/747.6 | 82746 | 19164 | 4.52 | 58874 | <0.001 |
| | 2 | Florida | 13 | 2019-01-01 to 2019-03-31 | (28.6N,85.5W)/1247.2 | 326625 | 94525 | 4.16 | 193772 | <0.001 |

# Correlation analysis of coastal states

As is shown in the figure, a high correlation was detected between 19 coastal states, although Delaware showed a negative correlation (Fig. 5). Interestingly, Mississippi and Texas, two high-risk states, showed a weak correlation. The reason may be as follows, since 2011, the incidence of ILI in these two states has been high and stable, while the incidence in other states were on a gradual upward trend.

# Time-series Analysis

Based on the result of spatio-temporal analysis, the HH cluster was identified mainly in Mississippi and Louisiana. In particular, Mississippi has been the HH cluster in recent years. It is necessary to predict the incidence of ILI in Mississippi by time-series analysis.

# SARIMA Model

Using raw training data from 1st week 2011 to 52nd week 2018, trend difference (d = 0) and seasonal difference (D = 1) were calculated. The Augmented Dickey-Fuller Test indicated the sequence was stationary ($t$=-3.98, $P$=

0.01). The ACF and PACF plots were used to estimate the parameter ranges of p, P and q, Q[19]. After checking ACF and PACF plots (Fig. 6), SARIMA(1, 0, 0)(1, 1, 0)$_{52}$ was the best fitted model with lowest AIC and BIC values, and this model passed the Ljung-Box Q Test (= 21.822,$P$ = 0.149), indicating it's a white noise sequence. All the parameter estimates were significant (Table 3).

**Table 3.** Comparison of candidate SARIMA models.

| el | Estimate | t | P | Ljung-Box Q Test | | AIC | BIC | RMSE | MAPE |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Statistics | P | | | | |
| MA (1,0,0)(1,1,0)$_{52}$ | - | - | - | 21.822 | 0.149 | 2235.530 | 2247.220 | 4.673 | 14.290 |
| | 0.886 | 36.768 | <0.001 | - | - | - | - | - | - |
| | -0.607 | 14.350 | <0.001 | - | - | - | - | - | - |
| MA (1,0,1)(1,1,0)$_{52}$ | - | - | - | 20.962 | 0.138 | 2235.110 | 2250.700 | 4.655 | 14.368 |
| | 0.865 | 28.837 | <0.001 | - | - | - | - | - | - |
| | 0.097 | 1.5410 | 0.065 | - | - | - | - | - | - |
| | -0.612 | -14.495 | <0.001 | - | - | - | - | - | - |
| MA (2,0,0)(1,1,0)$_{52}$ | - | - | - | 20.734 | 0.146 | 2201.100 | 2216.700 | 14.390 | 0.970 |
| | 0.957 | 18.254 | <0.001 | - | - | - | - | - | - |
| | -0.080 | -1.511 | 0.067 | - | - | - | - | - | - |
| | -0.612 | -14.495 | <0.001 | - | - | - | - | - | - |
| MA (2,0,1)(1,1,0)$_{52}$ | - | - | - | 18.552 | 0.183 | 2233.530 | 2253.012 | 4.636 | 14.602 |
| | 0.131 | 0.695 | 0.245 | - | - | - | - | - | - |
| | 0.653 | 3.744 | <0.001 | - | - | - | - | - | - |
| | 0.835 | 5.088 | <0.001 | - | - | - | - | - | - |
| | -0.605 | 14.157 | <0.001 | - | - | - | - | - | - |
| MA (2,0,2)(1,1,0)$_{52}$ | - | - | - | | | | | | |
| | -0.080 | -2.161 | 0.018 | 18.405 | 0.143 | 2233.480 | 2256.87 | 4.599 | 14.906 |
| | 0.825 | 26.101 | <0.001 | | | | | | |
| | 1.064 | 16.272 | <0.001 | | | | | | |
| | 0.064 | 0.994 | 0.163 | | | | | | |
| | -0.611 | -14.45 | <0.001 | | | | | | |

# Forecasting

The model SARIMA(1,0,0)(1,1,0)$_{52}$ forecasting effect was tested by comparing predicted values with observed values from 1st week 2019 to 29th week 2020. The results showed that all observed values were within 95%CI of predicted values, and the trend of predicted was basically consistent with the actual trend (Fig. 7). Then, forecasting the ILI incidence from 30th week 2020 to 52nd week 2021 by SARIMA. The forecast results show that from the 30th week 2020 to the 52nd week 2021, ILI will show a trend of high incidence in winter and spring, and low incidence in summer and autumn. The incidence of influenza will reach its peak in the 7th week 2021 (Table 4).

Table 4
Predictive value of ILI incidence (per 100,000)

| Year/week | Incidence | 95%CI | Year/week | Incidence | 95%CI |
|---|---|---|---|---|---|
| 2020/30 | 5.504 | -4.248-15.256 | 2021/16 | 11.536 | -8.684-31.755 |
| 2020/31 | 5.163 | -7.801-18.128 | 2021/17 | 10.344 | -9.876-30.563 |
| 2020/32 | 4.681 | -10.288-19.651 | 2021/18 | 8.225 | -11.995-28.444 |
| 2020/33 | 6.943 | -9.399-23.284 | 2021/19 | 7.892 | -12.327-28.112 |
| 2020/34 | 8.422 | -8.900-25.743 | 2021/20 | 7.633 | -12.587-27.853 |
| 2020/35 | 9.549 | -8.488-27.587 | 2021/21 | 6.467 | -13.753-26.687 |
| 2020/36 | 11.084 | -7.484-29.652 | 2021/22 | 7.606 | -12.613-27.826 |
| 2020/37 | 9.433 | -9.532-28.398 | 2021/23 | 5.670 | -14.550-25.890 |
| 2020/38 | 9.993 | -9.271-29.258 | 2021/24 | 5.411 | -14.809-25.63 |
| 2020/39 | 11.696 | -7.795-31.187 | 2021/25 | 5.784 | -14.436-26.004 |
| 2020/40 | 11.710 | -7.953-31.373 | 2021/26 | 5.468 | -14.751-25.688 |
| 2020/41 | 11.726 | -8.068-31.520 | 2021/27 | 4.498 | -15.721-24.718 |
| 2020/42 | 12.617 | -7.276-32.511 | 2021/28 | 4.990 | -15.230-25.21 |
| 2020/43 | 15.611 | -4.359-35.581 | 2021/29 | 4.672 | -15.548-24.892 |
| 2020/44 | 15.406 | -4.623-35.434 | 2021/30 | 4.823 | -15.851-25.496 |
| 2020/45 | 20.754 | 0.681–40.827 | 2021/31 | 5.010 | -16.006-26.025 |
| 2020/46 | 23.021 | 2.913–43.128 | 2021/32 | 4.910 | -16.364-26.184 |
| 2020/47 | 26.565 | 6.431–46.698 | 2021/33 | 6.056 | -15.414-27.526 |
| 2020/48 | 30.270 | 10.116–50.423 | 2021/34 | 8.693 | -12.926-30.313 |
| 2020/49 | 24.789 | 4.619–44.958 | 2021/35 | 9.995 | -11.739-31.728 |
| 2020/50 | 26.030 | 5.849–46.211 | 2021/36 | 11.233 | -10.587-33.054 |
| 2020/51 | 31.144 | 10.954–51.334 | 2021/37 | 9.925 | -11.962-31.812 |
| 2020/52 | 34.819 | 14.622–55.016 | 2021/38 | 10.378 | -11.560-32.316 |
| 2021/01 | 29.301 | 9.099–49.503 | 2021/39 | 11.596 | -10.380-33.573 |
| 2021/02 | 23.157 | 2.951–43.363 | 2021/40 | 11.996 | -10.010-34.003 |
| 2021/03 | 25.995 | 5.785–46.204 | 2021/41 | 11.853 | -10.176-33.883 |
| 2021/04 | 32.197 | 11.985–52.409 | 2021/42 | 12.471 | -9.577-34.518 |
| 2021/05 | 40.479 | 20.266–60.693 | 2021/43 | 17.086 | -4.975-39.147 |

| Year/week | Incidence | 95%CI | Year/week | Incidence | 95%CI |
|---|---|---|---|---|---|
| 2021/06 | 54.811 | 34.596–75.026 | 2021/44 | 15.498 | -6.574-37.569 |
| 2021/07 | 54.064 | 33.848–74.28 | 2021/45 | 23.641 | 1.562–45.720 |
| 2021/08 | 43.281 | 23.064–63.498 | 2021/46 | 26.943 | 4.858–49.028 |
| 2021/09 | 37.155 | 16.938–57.373 | 2021/47 | 33.262 | 11.172–55.352 |
| 2021/10 | 31.506 | 11.288–51.724 | 2021/48 | 36.767 | 14.674–58.86 |
| 2021/11 | 24.705 | 4.487–44.924 | 2021/49 | 29.557 | 7.461–51.654 |
| 2021/12 | 20.839 | 0.620-41.058 | 2021/50 | 30.704 | 8.606–52.802 |
| 2021/13 | 17.175 | -3.044-37.394 | 2021/51 | 36.928 | 14.828–59.027 |
| 2021/14 | 14.670 | -5.549-34.889 | 2021/52 | 40.739 | 18.638–62.84 |
| 2021/15 | 11.900 | -8.319-32.119 | | | |

# Discussion

This study explored the epidemiological characteristics of ILI, with particular emphasis on the discovery of high-risk states. In the descriptive analysis, we analyzed age characteristics, seasonal peaks and regional differences. The virus is most likely to cause severe symptoms and complications in children, elderly and immunocompromised individuals of all ages.[20] Because influenza viruses are transmitted through human contact, particularly geographic location and population density are potential factors of transmission and may indirectly lead to human death.[21] This study found high population density is more conducive to the spread of influenza viruses.

A gradual increase of incidence was observed, particularly during the 2018 and 2019 influenza seasons, with a sharp increase in the incidence of ILI. The local Moran's I detected the clusters based on the administrative divisions.[16] The high incidence of influenza was mainly concentrated in the states of Louisiana, Virginia and Mississippi, and the Local spatial autocorrelation analysis revealed the HH cluster was mainly located in Louisiana and Mississippi. This means that if the influenza incidence is high in Louisiana and Mississippi, the neighboring states will also have higher influenza incidence rates. The temporal scan statistic showed the high incidence time mainly occurs during January to March. This finding was confirmed by other studies.[6, 22]

Transmission of influenza varies across seasons and geographical areas in the United States. The obvious temporal clusters fell during the winter and spring, which was in accordance with the seasonality of the respiratory disease.[23] Factors that affect the spread of influenza include changes in weather variables such as temperature, humidity, or rainfall. In an annual winter epidemic in the United States, the spread of influenza increases during periods of high precipitation and humidity.[24] Most parts of the United States belong to the temperate and subtropical climate. The continental climate zone in the central plain is characterized by a continental climate with cold winters. The average temperature of winter in temperate zone is between 0°C and 20°C, and the minimum temperature in some areas drops to − 40°C. The appropriate temperature range for influenza virus transmission may explain in part the common winter epidemics in the central region.[25] In Tropical south-east Asia, environmental temperature and humidity may play an important role in virus

transmission and virus survival, depending on the size of virus particles.[25] The correlation analysis of coastal states showed a high correlation. Appropriate temperature and humidity may be the reason for the high incidence in the southeast.

Based on past data, Mississippi were the high-risk areas in influenza season, and should be a priority surveillance area for influenza. A time series model was used to predict the future incidence of Mississippi. Time series analysis has the advantage of predicting the incidence and doesn't need to pay attention to the specific risk factors. It is characterized by the number of patients in the past and responds by predicting the number of patients in the future. The model was established by ACF and PACF plots. According to the BIC, residual white noise test and fitting effect, the regression model SARIMA$(1, 0, 0)(1, 1, 0)_{52}$ with statistical significance was obtained.[4] The prediction showed that the 95% confidence interval of the predicted ILI incidence almost contains the observed value, which supports that the SARIMA model is effective in the prevention of ILI. Then, we used this model to forecast the ILI incidence from 30th week 2020 to 52nd week 2021, the results demonstrated that ILI incidence will begin to increase in 45th week 2020 and peak in 6th week 2021, the distribution is similar to the previous years.

# Limitation

Our results also have some limitations. First, our study included all cases reported by the United States outpatient influenza-like illness surveillance network, but there may still be undetected cases. This may affect the accuracy of the use of ILI forecasts to estimate influenza activity. Second, influence factors of influenza activity were not conducted in-depth studied in this study. In the next step, we need to collect data for risk factor analysis.

## Conclusion

The observed of ILI showed a gradually upward trend from 2011 to 2020, especially in the last two years. Inadequate Influenza Vaccine Supplementation, the emergence of influenza A (H1N1)pdm09 and the expansion of influenza surveillance efforts may be the main reasons for the dramatic changes in influenza outbreaks and spatio-temporal patterns.[22] States of relatively high risk for influenza have been identified in United States. In order to explain the spatio-temporal pattern of influenza, it is necessary to conduct more future studies on risk factors at the national and local levels.

## Declarations

# Ethics approval and consent to participate

The data used or obtained in this study is a publicly available data, it can be find with https://www.cdc.gov/flu/weekly/fluactivitysurv.htm

# Consent for publication

Not applicable

# Availability of data and materials

The data used or obtained in this study is a publicly available data, which can be available in the United States Outpatient Influenza-like Illness Surveillance Network repository (https://www.cdc.gov/flu/symptoms/testing.htm.)

# Competing interests

The authors declare that they have no competing interests.

# Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

# Authors' contributions

The manuscript was completed mainly by Zhijuan song. Junzhe bao and Huili zhu put forward some suggestions for the revision of this study. Xuezhong shi provided guidance on the writing of the manuscript and corrected errors. Xiaocan jia and Yongli yang read the manuscript and revised it. All authors have full access to all aspects of the research and writing process, and taking final responsibility for the paper.

# Acknowledgments

We thank Xuezhong Shi, Yongli Yang, Xiaocan Jia, Junzhe Bao and Huili Zhu for helpful discussions.

The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Health Resources and Services Administration or HHS.

No financial disclosures were reported by the authors of this paper.

## References

1. Tokars JI, Olsen SJ, Reed C: **Seasonal Incidence of Symptomatic Influenza in the United States**. *Clin Infect Dis* 2018, **66**(10):1511–1518.
2. Lu J, Meyer S: **Forecasting Flu Activity in the United States: Benchmarking an Endemic-Epidemic Beta Model**. *Int J Environ Res Public Health* 2020, **17**(4).
3. Iuliano AD, Roguski KM, Chang HH, Muscatello DJ, Palekar R, Tempia S, Cohen C, Gran JM, Schanzer D, Cowling BJ *et al*: **Estimates of global seasonal influenza-associated respiratory mortality: a modelling study**. *The Lancet* 2018, **391**(10127):1285–1300.
4. Rao X, Chen Z, Dong H, Zhu C, Yan Y: **Epidemiology of influenza in hospitalized children with respiratory tract infection in Suzhou area from 2016 to 2019**. *J Med Virol* 2020.

5. **Diagnosing Flu** [https://www.cdc.gov/flu/symptoms/testing.htm]
6. Lu FS, Hattab MW, Clemente CL, Biggerstaff M, Santillana M: **Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches**. *Nat Commun* 2019, **10**(1):147.
7. Salimi M, Jesri N, Javanbakht M, Farahani LZ, Shirzadi MR, Saghafipour A: **Spatio-temporal distribution analysis of zoonotic cutaneous leishmaniasis in Qom Province, Iran**. *Journal of Parasitic Diseases* 2018, **42**(4):570–576.
8. Alimohamadi Y, Zahraei SM, Karami M, Yaseri M, Lotfizad M, Holakouie-Naieni K: **Spatio-temporal analysis of Pertussis using geographic information system among Iranian population during 2012–2018**. *Med J Islam Repub Iran* 2020, **34**:22.
9. Barnea O, Huppert A, Katriel G, Stone L: **Spatio-temporal synchrony of influenza in cities across Israel: the "Israel is one city" hypothesis**. *PLoS One* 2014, **9**(3):e91909.
10. Petukhova T, Ojkic D, McEwen B, Deardon R, Poljak Z: **Assessment of autoregressive integrated moving average (ARIMA), generalized linear autoregressive moving average (GLARMA), and random forest (RF) time series regression models for predicting influenza A virus frequency in swine in Ontario, Canada**. *PLoS One* 2018, **13**(6):e0198313.
11. He D, Lui R, Wang L, Tse CK, Yang L, Stone L: **Global Spatio-temporal Patterns of Influenza in the Post-pandemic Era**. *Sci Rep* 2015, **5**:11013.
12. Dong W, Yang K, Xu QL, Yang YL: **A Predictive Risk Model for A(H7N9) Human Infections Based on Spatial-Temporal Autocorrelation and Risk Factors: China, 2013–2014**. *Int J Environ Res Public Health* 2015, **12**(12):15204–15221.
13. Gu H, Fan W, Liu K, Qin S, Li X, Jiang J, Chen E, Zhou Y, Jiang Q: **Spatio-temporal variations of typhoid and paratyphoid fevers in Zhejiang Province, China from 2005 to 2015**. *Sci Rep* 2017, **7**(1):5780.
14. Fuentes-Vallejo M: **Space and space-time distributions of dengue in a hyper-endemic urban space: the case of Girardot, Colombia**. *BMC Infect Dis* 2017, **17**(1):512.
15. Zhu B, Liu JL, Fu Y, Zhang B, Mao Y: **Spatio-Temporal Epidemiology of Viral Hepatitis in China (2003–2015): Implications for Prevention and Control Policies**. *Int J Env Res Pub He* 2018, **15**(4).
16. Mao Y, He R, Zhu B, Liu J, Zhang N: **Notifiable Respiratory Infectious Diseases in China: A Spatial-Temporal Epidemiology Analysis**. *Int J Environ Res Public Health* 2020, **17**(7).
17. Khudyakov YE, Liu Y, Wang X, Liu Y, Sun D, Ding S, Zhang B, Du Z, Xue F: **Detecting Spatial-Temporal Clusters of HFMD from 2007 to 2011 in Shandong Province, China**. *PLoS ONE* 2013, **8**(5).
18. Kulldorff M, Athas WF, Feurer EJ, Miller BA, Key CR: **Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico**. *Am J Public Health* 1998, **88**(9):1377–1380.
19. Cong J, Ren M, Xie S, Wang P: **Predicting Seasonal Influenza Based on SARIMA Model, in Mainland China from 2005 to 2018**. *Int J Environ Res Public Health* 2019, **16**(23).
20. Keilman LJ: **Seasonal Influenza (Flu)**. *Nurs Clin North Am* 2019, **54**(2):227–243.
21. Chandra S, Kassens-Noor E, Kuljanin G, Vertalka J: **A geographic analysis of population density thresholds in the influenza pandemic of 1918-19**. *Int J Health Geogr* 2013, **12**:9.
22. Zhang Y, Wang X, Li Y, Ma J: **Spatiotemporal Analysis of Influenza in China, 2005–2018**. *Scientific Reports* 2019, **9**(1).

23. Chow EJ, Doyle JD, Uyeki TM: **Influenza virus-related critical illness: prevention, diagnosis, treatment**. *Crit Care* 2019, **23**(1):214.
24. Gomez-Barroso D, Leon-Gomez I, Delgado-Sanz C, Larrauri A: **Climatic Factors and Influenza Transmission, Spain, 2010–2015**. *Int J Environ Res Public Health* 2017, **14**(12).
25. Dave K, Lee PC: **Global Geographical and Temporal Patterns of Seasonal Influenza and Associated Climatic Factors**. *Epidemiol Rev* 2019, **41**(1):51–68.

# Figures

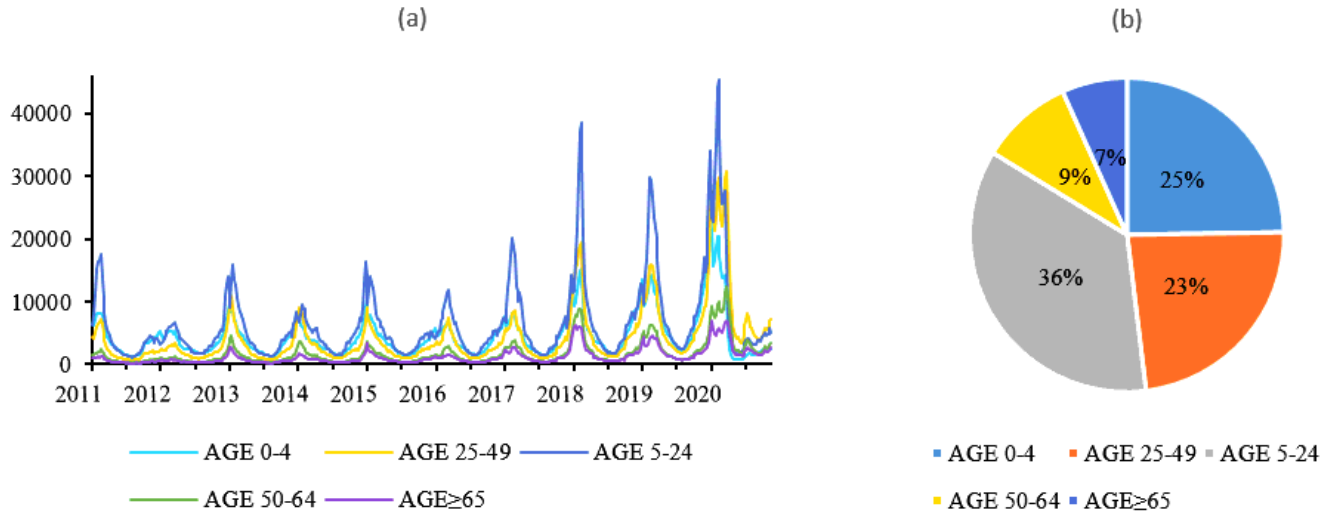



**Figure 1**

Epidemiological characteristics of influenza and ILI in the United States from 1st week 2011 to 29th week 2020. (a) Weekly ILI cases of different age (b) Age distribution of ILI

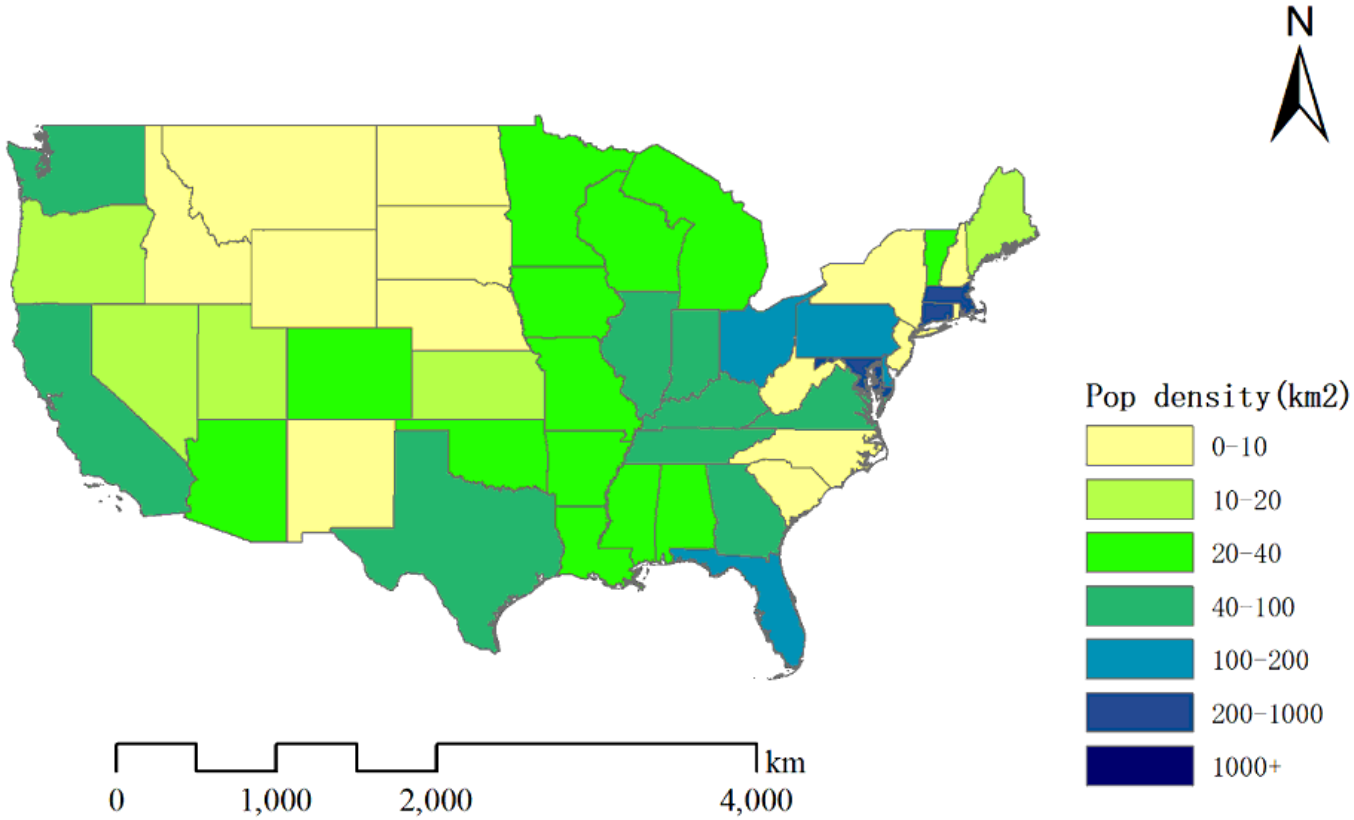
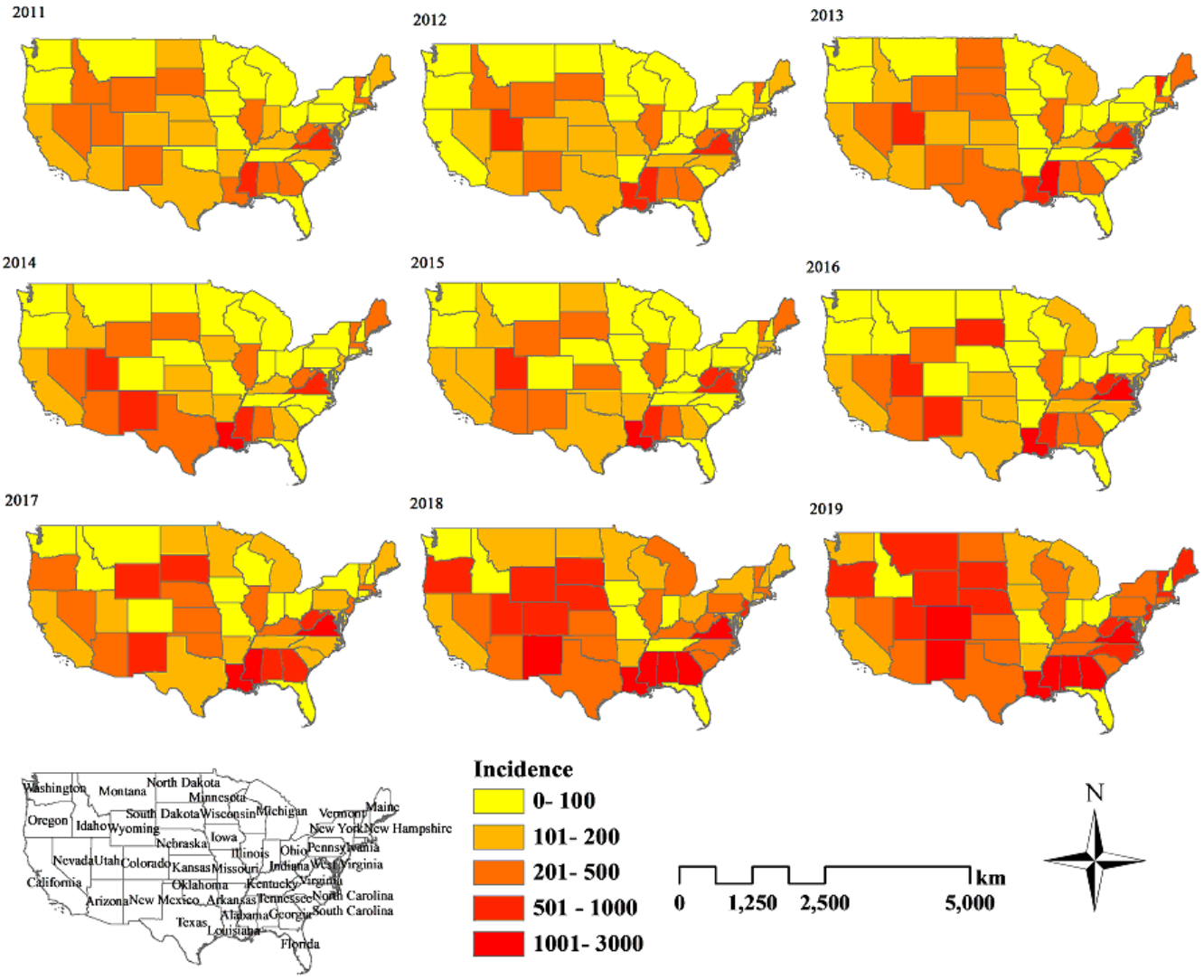


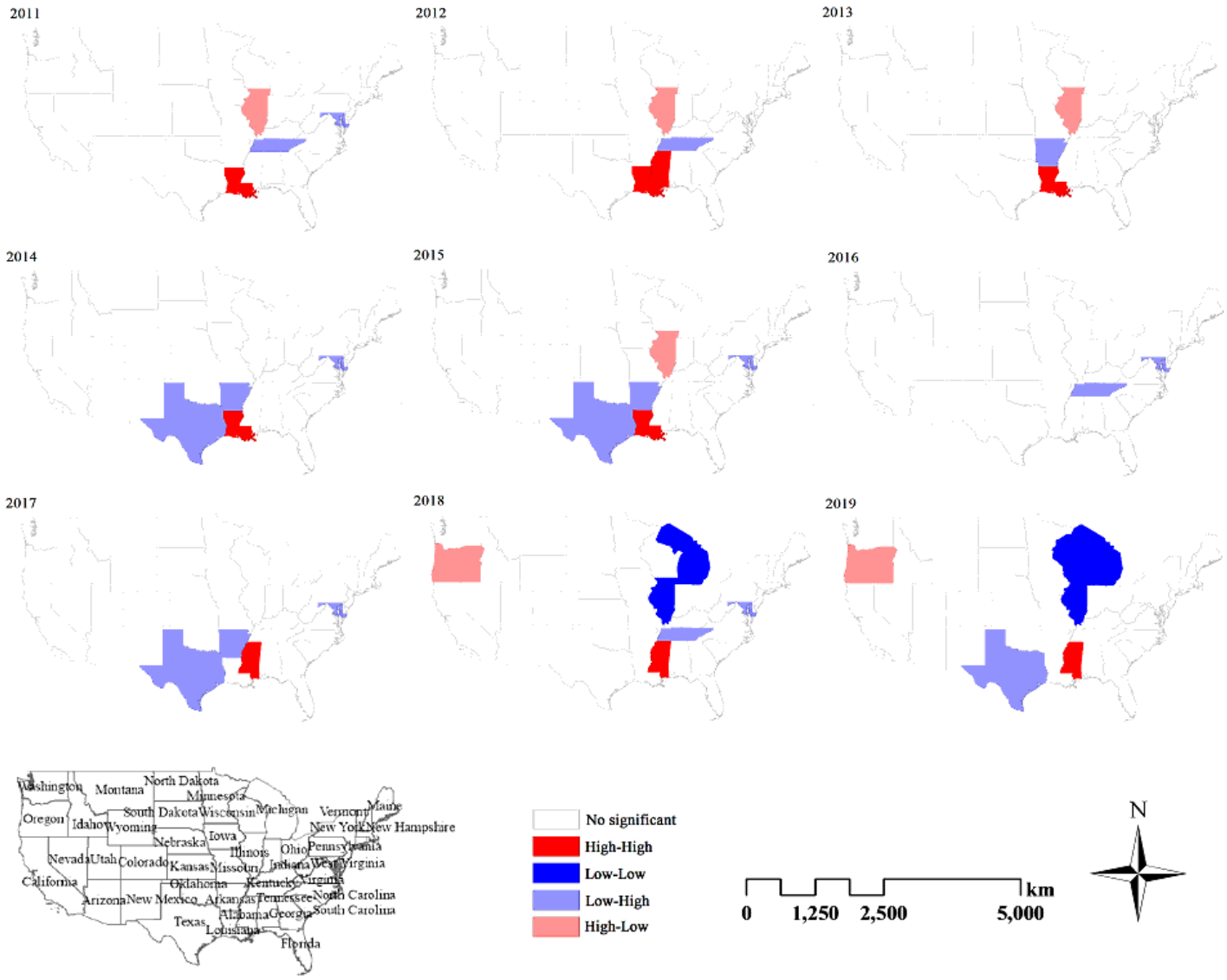

**Figure 2**

Map of population density by state in the United States. The map was created by ArcGIS software.



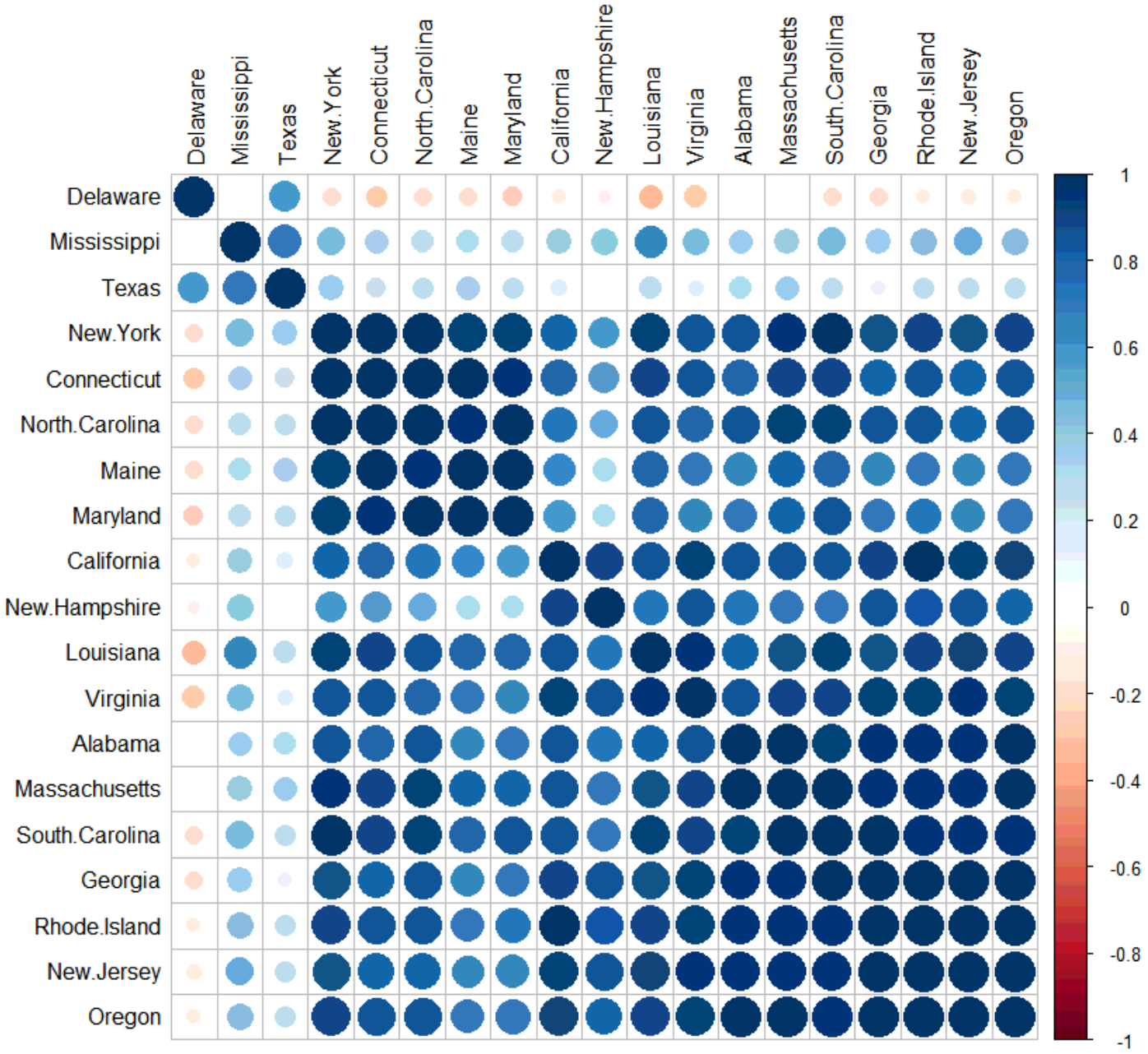

## Figure 3

Maps of ILI incidence for each state in the United States from 2011 to 2019. These maps were created by ArcGIS software.



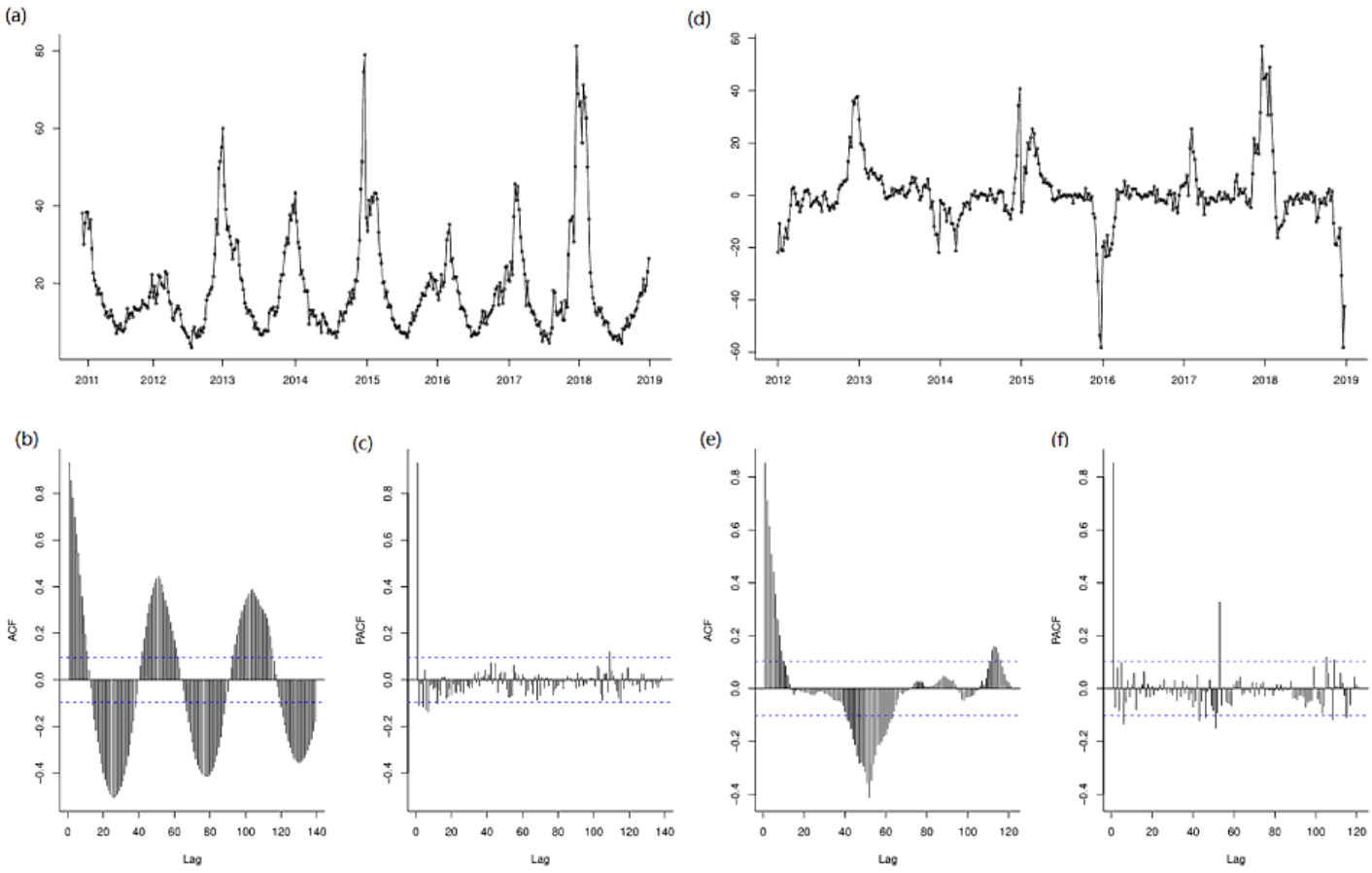

## Figure 4

Maps of Local spatial autocorrelation cluster of ILI for each state in the United States from 2011 to 2019. These maps were created by GeoDa software.




## Figure 5

Heatmap of ILI incidence correlations between 19 coastal states in the study period. These Heatmap were created by Rstudio.




## Figure 6

The time diagram, ACF and PACF graphs for estimating the parameter:(a) The time diagram of ILI incidence after one-order seasonal difference data, (b) the ACF graph of the raw data(d=0, D=0), (c) the PACF graph of the raw data(d=0, D=0), (d) the time diagram of ILI incidence a, (e) the ACF graph of one-order seasonal difference data(d=0 and D=1),(f) the PACF graph of one-order seasonal difference data(d=0 and D=1)

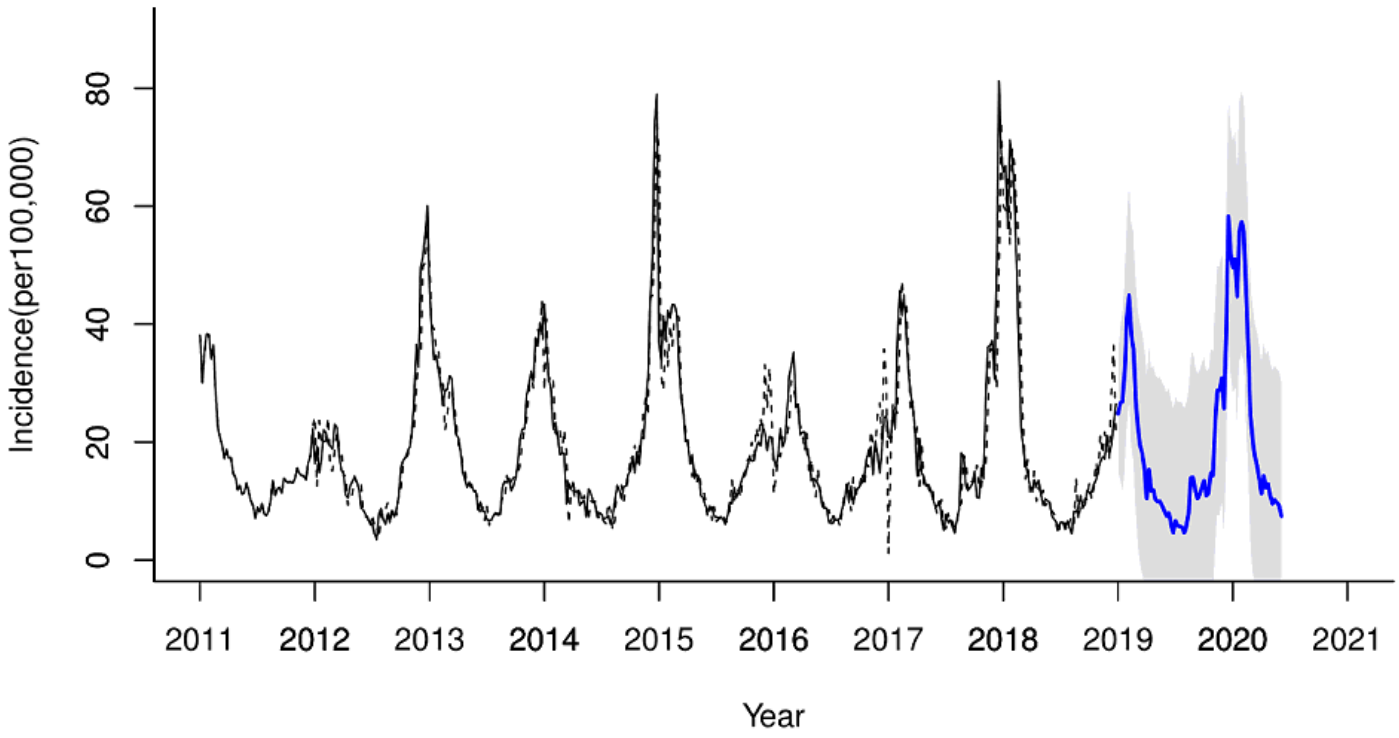



**Figure 7**

Comparison of actual and predicted incidence of ILI in the United States.