

1 **Spatial transmission network construction of influenza-like illness using Dynamic**

2 **Bayesian Network and Vector-Autoregressive Moving Average Model**

3 *Jianqing Qiu¹, Huimin Wang¹, Tao Zhang^{1*}, Changhong Yang^{2*}*

4

5 Jianqing Qiu¹, M.M., qiujianqing0219@163.com

6 Huimin Wang¹, B.M., 2773201051@qq.com

7 Tao Zhang^{1*}, D. M., scdxzhangtao@163.com

8 Changhong Yang^{2*}, Professor, changhong_yang@163.com

9

10 1. Department of Epidemiology and Health Statistics, West China School of Public

11 Health and West China Fourth Hospital, Sichuan University, Sichuan, China

12 2. Sichuan Center for Disease Control and Prevention

13

14 ***Corresponding author:**

15 Tao Zhang

16 Renmin South Road 3rd Section NO.16

17 Chengdu 610041, Sichuan

18 Department of Epidemiology and Health Statistics

19 West China School of Public Health and West China Fourth Hospital, Sichuan

20 University

21 E-mail: scdxzhangtao@163.com

22 Tel number: +8615908141271

23

24 Changhong Yang

25 Zhongxue Road NO.6

26 Chengdu 610041, Sichuan

27 Sichuan Center for Disease Control and Prevention

28 E-mail: changhong_yang@163.com

29 Tel number: +8618981958752

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46 **Abstract:**

47 **Background:** Influenza is an acute respiratory infection caused by an influenza virus,
48 and the primary intervention strategy is seasonal vaccine. Due to various influenza
49 strains and their rapid mutation each year, how to recognize the key population and
50 timing of the vaccination becomes essential. Considering the importance of finding
51 possible spreading directions and effects of influenza between cities for department of
52 influenza prevention, the construction of influenza transmission network becomes
53 meaningful.

54 **Methods:** 21 cities in Sichuan province were divided into different learning
55 communities according to whether they were adjacent to each other or not. In each

56 community, the first-order conditional dependencies approximation algorithm was
57 performed to learn the possible structure of the time-lagged correlations between
58 different time series vectors of the ILI estimated weekly number, and the vector
59 autoregressive moving average models were performed for learning the lag orders and
60 parameters of the time-lagged correlations between different time series vectors in each
61 community.

62 **Results:** It detected a number of significant time-lagged correlations between cities in
63 Sichuan province using two models, and the lag was from 1 week to 3 weeks. The
64 parameters indicating the suspected propagation relationship were between -0.90 and
65 0.75, and the proportion of the negative values in parameters increased with time.
66 Furthermore, the spreading routes learning from two models were almost in accordance
67 with the traffic network of Sichuan province.

68 **Conclusions:** This study proposed an innovative framework for exploring the
69 potentially stable transmission routes between different regions and measuring specific
70 size of the transmission effect. It could be used for the infectious disease key area
71 confirmation by considering their adjacent areas' incidence and the transmission
72 relationship.

73

74 **Keywords:** Influenza; Spatial transmission network; Dynamic Bayesian Network;
75 Vector Autoregressive Moving Average Model

76

77 **1 Background**

78 Influenza is an acute respiratory infection caused by an influenza virus. According to
79 the characteristics of pathogens, it can be divided into four types: Type A, Type B, Type
80 C, and Type D¹. Due to antigenic drift of the glycoprotein hemagglutinin (HA)—the
81 reassortment of gene segments between two different viruses, the influenza virus has
82 caused several worldwide pandemics. According to the globally estimated disease
83 burden attributed to the influenza extrapolated from developed countries, it is estimated
84 there are about 1 billion seasonal influenza cases per year on average, among which
85 there are three million serious cases, resulting in 250,000 to 500,000 deaths².

86 As for influenza, also a representative infectious disease, the primary intervention
87 strategy is seasonal vaccine^{3,4}. However, the antigenic drift often makes the vaccine-
88 induced immunity wane over the course of a season⁴. Specifically, protective antibodies
89 are usually produced after 2-4 weeks of influenza vaccination, and antibody titers
90 decreases with time, totally decaying to the level not enough to protect the recipients
91 after 6-8 months⁵. Due to various influenza strains and their rapid mutation each year,
92 the WHO rescreens different strains and recommends variously corresponding vaccines
93 for the prevention. Therefore, the development and implementation of influenza
94 vaccines becomes a yearly routine strategy, bringing fairly a huge burden to low-to-
95 middle income countries. Considering the huge cost of the a series of prevention
96 strategies on influenza, it is necessary to learn how to efficiently use those limited
97 resources on prevention and control of influenza⁶. One acceptedly related topic is to
98 recognize the key population and timing of the vaccination, so the study on spatial-

99 temporal epidemic trend of infectious disease including influenza is considerably
100 valuable.

101 For department of influenza prevention and control such as Disease Control and
102 Prevention Center (CDC), among the research about spatial-temporal epidemic trend
103 of infectious disease, one essential topic they concern about was the influenza spreading
104 directions between variously adjacent regions and how much impact an influenza
105 situation of a region has on an adjacent region, and those kind of ‘influenza transmission
106 network construction’ research can guide them to better allocate prevention and control
107 resource. Therefore, the topic on construction of influenza transmission network
108 becomes popular and meaningful.

109 For solving the problem of constructing possible transmission network, traditional
110 epidemic modeling on complex networks⁷ was utilized for prediction of the epidemic
111 trend by constructing possible transmission network using population contact data in a
112 mathematical way. In addition, another study has observed and utilized birds migration
113 network to predict the developing trajectory of avian influenza⁸. Those methods need
114 personal contact data including person mobility data, traffic data and avian mobility
115 data. However, there will be obstacle on the availability of those data, and collecting
116 those data will cost too much human work or time, and usually it is not on time for
117 making rapid prevention strategies of influenza with those research.

118 Based on the aim to construct a transmission network of influenza and referring to
119 previous studies, we proposed a new concept calling ‘spatial transmission route’ to

120 display the transmission directions of influenza and to measure the specific size of those
121 transmission effect. Particularly, we defined ‘spatial transmission route’ as time-lagged
122 association between different regions’ epidemiological indexes (absolute case number,
123 incidence and so on). This method does not need personal contact data, and only
124 depends on surveillance data. The concept of “spatio-temporal association” is to
125 explore the transmission network themselves using real-world surveillance data in
126 statistical framework.

127 In this study, we selected Sichuan province for research because Sichuan was a
128 relatively confined area due to its mountainous landscape in southwestern China. We
129 assumed that in this relatively confined area the spreading direction and spreading
130 impacts were relatively stable. By exploring the possible transmission network in
131 Sichuan, we can quantify and qualify the transmission effects between different regions.
132 The findings will suggest the extent of being infected in some regions by neighboring
133 regions in an influenza epidemic, and guide an idea for the department to make
134 prevention and control strategies by finding key regions of prevention.

135 The remainder of this paper is organized as follows: In Section 2, we introduce the
136 concept of possible transmission route, the dynamic Bayesian network (DBN), and
137 Vector-Autoregressive Moving Average Model (VARMA). In Section 3, we describe
138 the time, region and population distribution characteristics of influenza-like cases
139 monitored by 30 sentinel hospitals in 21 cities and municipalities in Sichuan province
140 from 2010 to 2016 and exhibit possible transmission routes of infectious disease in
141 Sichuan province by means of DBN and VARMA.

142

143 **2 Methods**

144 **2.1 Data acquisition**

145 Our data was composed of the sentinel surveillance data on influenza-like illness of
146 Sichuan province covering 30 sentinel surveillance hospitals in 21 cities and Anyue,
147 Santai County across entire province from 2010 to 2016. This data was obtained from
148 the Sichuan Provincial Center for Disease Control and Prevention. As shown in
149 Additional file 1, there was at least one sentinel hospital in a city of Sichuan. According
150 to the report requirement, the medical staffs of monitoring clinics in sentinel hospitals
151 recorded the number of influenza-like illness (ILI) and the total number of outpatients
152 cases in each age group in each department every day, and uploaded the data to ‘China
153 Influenza Surveillance Information System’ before 24 o'clock every Monday. For this
154 study, the definition of influenza-like illness (ILI) referring to WHO⁹ was as follows: a
155 case measured fever of $\geq 38\text{ C}^\circ$ and cough; with onset within the last 10 days. Besides,
156 to estimate the absolute ILI case number in city j , we collected the data of yearly number
157 of medical outpatients in each city of Sichuan from ‘Sichuan Health Statistics Yearbook
158 2010~2012’ and ‘Sichuan Health and Family Planning Statistical Yearbook
159 2013~2016’.

160 **2.2 The definition of transmission route**

161 As mentioned before, Alonso WJ¹⁰ used Fourier decomposition to describe the
162 amplitude and timing of annual and semiannual epidemic cycles of influenza B in Brazil,

163 and compared seasonal parameters across latitudes. He found a seasonal southward
 164 traveling wave of influenza across Brazil originating from equatorial and low
 165 population regions in March–April and moving towards temperate and highly popular
 166 regions over a 3-month period. The result illuminated that this kind of spatio-temporal
 167 relationship of infectious diseases may be the transmission traveling route. Also in Paul
 168 and Held’s study¹¹, they proposed a random effect model, called epidemic-endemic
 169 model considering the transmission effects of neighboring units as follows. The
 170 component of $\phi_i \sum_{j \neq i} w_{ji} y_{j,t-1}$ modeled the transmission effect of j area on i area, and used
 171 w_{ij} to weight those effects according the corresponding traffic adjacent degree.

$$172 \quad \zeta_{i,t} = \lambda_i y_{i,t-1} + \phi_i \sum_{j \neq i} w_{ji} y_{j,t-1}$$

173 In this study, referring to the finding of Alonso and the framework of Held, we defined
 174 the time-lagged correlations between two cities’ influenza case number vector as the
 175 suspected spatial transmission route of influenza among cities in Sichuan province. In
 176 this study, for ILI case number time series vectors $X_{1,L}, X_i, X_j, L$ in city 1, \dots , i ,
 177 j , \dots , if there existed a statistically significant correlation between $X_i(t)$ and
 178 $X_j(t-1)$, we considered there existed a suspected transmission route from city j to city
 179 i , and the transmission effect emerged at one time-lagged unit.

180 Considering that our aim was to model time-lagged associations between different time
 181 series vectors, dynamic Bayesian networks and Vector-Autoregressive Moving Average
 182 Model were ideal tools for solving those problems. The introduction of those two tools
 183 were as follows.

184 **2.3 Dynamic Bayesian network**

185 Bayesian Network is a directed acyclic graph (DAG) defining the conditional
186 probabilistic dependencies between a set of random variables¹². Due to its superiority
187 on modelling time-lagged correlation between various vectors, we regarded it as an
188 ideal tool to solve the problem of influenza transmission route. As the following
189 formula shows, it is composed of two components: V representing the set of nodes,
190 among which each node $v_i \in V$ corresponds to a random variable X_i ; A representing the
191 set of arrows, showing the probabilistic dependencies between variables.

192
$$G = (V, A)$$

193 Dynamic Bayesian Network (DBN) is an extension of the Bayesian network model for
194 identifying dynamic relationships between nodes. It introduces the concept of timing
195 based on a static Bayesian network. Unlike static Bayesian networks, each variable in
196 a dynamic Bayesian network is represented several nodes across time points. The result
197 of DBN rests on three assumptions: (1) The stochastic process X is first-order
198 Markovian; (2) The random variables $X(t) = (X_1(t), K, X_i(t), K, X_k(t))$ are
199 conditionally independent given the random variables $X(t-1)$ at the previous time
200 $t-1$; (3) The temporal profile $(X_i(1), K, X_i(n))$ of any variable X_i cannot be written as
201 a linear combination of the other profiles $(X_j(1), K, X_j(n))$. The probability
202 distribution of the network is as follows:

203
$$P(g_{1,t-1}, g_{2,t-1}, L, g_{n,t-1}, g_{1,t}, g_{2,t}, L, g_{n,t}) = \prod_{i=1}^n P(g_{i,t} | Pa(g_{i,t}), \theta)$$

204 The application of DBN in the field of systems biology has been matured. Due to the
205 superiority of network structure learning, it is often used to characterize the gene
206 regulatory network by characterizing the correlation of multiple gene expression data.
207 It is revealed as the best method for complex biological regulatory networks
208 modelling^{13,14}. Because the surveillance data of infectious disease also has four similar
209 characteristics: high noise, nonlinear correlation, small sample and latent variables as
210 the gene expression data¹⁵, the surveillance data of infectious disease can also be fitted
211 to the DBN model theoretically, and it has been initially proved a good fit handling four
212 challenges mentioned in infectious disease surveillance data¹⁵, and proved a good
213 function in infectious disease prediction model^{16,17}.

214 Classic dynamic Bayesian network algorithms include the Least Absolute Shrinkage
215 and Selection Operator or LASSO¹⁸, James–Stein Shrinkage¹⁹, First-Order Conditional
216 Dependencies Approximation²⁰ and Modular Networks²¹. Study²⁰ shows that Low-
217 Order Conditional Dependencies Approximation (Also called G1DBN) has a better
218 precision and sensitivity attributing to dimension deduction. Therefore, in this study,
219 we intend to adopt Low-Order Conditional Dependencies Approximation for structure
220 learning.

221

222 **2.4 Vector-Autoregressive Moving Average Model**

223 Vector-autoregressive model (VAR) was another model for processing dynamic
224 relationship between simultaneously multiple time series²² based on Granger causality,

225 and it was firstly proposed by Christopher Sims⁶. The basic formula of VAR(p) was:

226
$$X(t) = A_1 X(t-1) + L + A_l X(t-l) + A_p X(t-p) + B + \varepsilon(t)$$

227
$$X(t) = \begin{bmatrix} X_1(t) \\ L \\ X_i(t) \\ L \\ X_j(t) \\ L \\ X_k(t) \end{bmatrix}, A_l = \begin{bmatrix} a_{l,11} L & a_{l,1i} L & a_{l,1j} L & a_{l,1k} \\ O & & & N \\ & a_{l,ii} L & a_{l,ij} \\ & M & M \\ & a_{l,ji} L & a_{l,jj} \\ & N & & O \\ a_{l,k1} L & a_{l,ki} L & a_{l,kj} L & a_{l,kk} \end{bmatrix}, \varepsilon(t) = \begin{bmatrix} \varepsilon_1(t) \\ L \\ \varepsilon_i(t) \\ L \\ \varepsilon_j(t) \\ L \\ \varepsilon_k(t) \end{bmatrix}$$

228 And the basic formula of VARMA(p, q) was:

229
$$X(t) = A_0 + A_1 X(t-1) + L + A_l X(t-l) + L + A_p X(t-p) + \varepsilon(t) - B_1 \varepsilon(t-1) - L - B_l \varepsilon(t-l) - L - B_p \varepsilon(t-q)$$

230
$$X(t) = \begin{bmatrix} X_1(t) \\ L \\ X_i(t) \\ L \\ X_j(t) \\ L \\ X_k(t) \end{bmatrix}, A_l = \begin{bmatrix} a_{l,11} L & a_{l,1i} L & a_{l,1j} L & a_{l,1k} \\ O & & & N \\ & a_{l,ii} L & a_{l,ij} \\ & M & M \\ & a_{l,ji} L & a_{l,jj} \\ & N & & O \\ a_{l,k1} L & a_{l,ki} L & a_{l,kj} L & a_{l,kk} \end{bmatrix}, \varepsilon(t) = \begin{bmatrix} \varepsilon_1(t) \\ L \\ \varepsilon_i(t) \\ L \\ \varepsilon_j(t) \\ L \\ \varepsilon_k(t) \end{bmatrix},$$

231
$$B_l = \begin{bmatrix} b_{l,11} L & b_{l,1i} L & b_{l,1j} L & b_{l,1k} \\ O & & & N \\ & b_{l,ii} L & b_{l,ij} \\ & M & M \\ & b_{l,ji} L & b_{l,jj} \\ & N & & O \\ b_{l,k1} L & b_{l,ki} L & b_{l,kj} L & b_{l,kk} \end{bmatrix}$$

232 The addition of the moving average term $-B_1 \varepsilon(t-1) - L - B_l \varepsilon(t-l) - L - B_p \varepsilon(t-q)$

233 improved the fit function by extracting non-stationary deterministic information from

234 the original series. Since the DBN and the VARMA model could be used to characterize

235 the time-lagged correlation between time series vectors, and their mathematical
 236 principles were similar, the following relationship existed between them: if there
 237 existed an arrow from $X_i(t-1)$ to $X_j(t)$ in the result of structure learning, then
 238 $\Lambda a_{l,ij} \neq 0, l \in (1, p)$ in the extended cross matrix of the corresponding VARMA(p, q)
 239 in parameter learning part.

240

241 **2.5 Data processing**

242 Excel 365 was used for data management, and R 3.5.0 was utilized for data cleaning
 243 and modeling. The main indicator describing the possible incidence of influenza in city
 244 j was the percentage of Influenza-like Illness (ILI%), a commonly used indicator for
 245 national influenza surveillance. As was shown in formula (1), it referred to the
 246 percentage of ILI case number occupying in the number of visits to medical
 247 organizations in hospital i and week r .

$$248 \quad ILI\%(i, r) = \frac{ILI \text{ case number}(i, r)}{Visits \text{ to medical organizations}(i, r)} \quad (1)$$

249 In the part of DBN network generation, on one hand, considering the fact that it was
 250 the absolute ILI case number in city j and week r making sense in the influenza
 251 spreading to other adjacent cities in following weeks rather than the relative number
 252 ILI%, we firstly estimated the $ILI \text{ case number}(j, r)$ in city j and week r using the
 253 $ILI\%(i, r)$ in surveillance hospital i and week r , and
 254 $Visits \text{ to medical organizations}(j, s)$ in city j and year s , and the formulas calculation

255 was in formula (2):

$$\begin{aligned} 256 \quad & ILI \text{ case number}(j, r) \approx \frac{ILI \text{ case number}(i, r)}{Visits \text{ to medical organizations}(i, r)} \quad (2) \\ 257 \quad & \times Visits \text{ to medical organizations}(j, r) \\ 258 \quad & \approx \frac{ILI \text{ case number}(i, r)}{Visits \text{ to medical organizations}(i, r)} \\ 259 \quad & \times \frac{Visits \text{ to medical organizations}(j, r)}{Visits \text{ to medical organizations}(i, s)} \\ 260 \quad & \times Visits \text{ to medical organizations}(j, s) \\ 261 \quad & \approx ILI \text{ case number}(i, r) \\ 262 \quad & \times \frac{Visits \text{ to medical organizations}(j, s)}{Visits \text{ to medical organizations}(i, s)} \end{aligned}$$

263 Then we could regard $ILI \text{ case number}(j, r)$ time series as a vector, and our task was
264 to model the time-lagged correlation between every vector $ILI \text{ case number}(j, r)$.
265 On other hand, considering the interpretability of the result, for example, we assumed
266 that the influenza spreading mainly happened between adjacent cities, and non-adjacent
267 influenza spreading within a province did not happen. Therefore, we put every
268 geographically adjacent city as one basic learning unit, and the specific list was in
269 Additional file 2. Structure learning and parameter learning were used to characterize
270 the time-lagged correlations, so-called transmission routes between the vectors in each
271 unit. In the part of structure learning, we adopted the first-order condition
272 approximation algorithm, also called G1DBN to generate the DBN characterizing the

273 influenza transmission routes between every city in a basic unit quantificationally, and
274 the DBN was generated using G1DBN package, available for the R programming
275 environment. In the part of parameter learning, we used the VARMA model to measure
276 the effect of significantly time-lagged correlation in structure learning, and the VARMA
277 model was fitted using MTS package, also available for the R programming
278 environment. Finally, we put every influenza transmission route in each unit together
279 to form a transmission network.

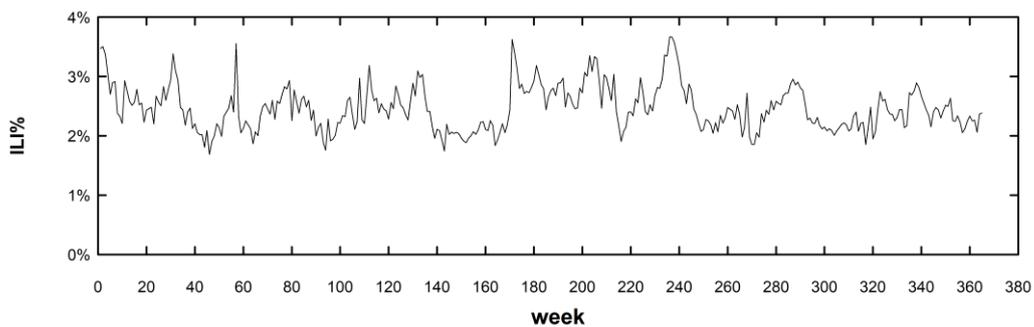
280 The VARMA(p, q) model was fitted in each unit in parameter learning. For a unit
281 composed of A, B and C, if there was an arrow ‘ \rightarrow ’ between weekly estimated vectors
282 from A to B in the part of structure learning, it meant that there was a significantly
283 meaningful transmission route from A to B, otherwise no significantly transmission
284 route between two cities was detected. Accordingly, we fitted VARMA model in the
285 community of A, B and C cities with fixing the coefficients between A and C, B and C
286 in each time-lagged order to zero. Then we utilized VARMA model for parameter
287 learning, and the procedure of fitting VARMA model was specified in Additional file 3,
288 the results of VARMA implied two essential information: Firstly, the order selection
289 result displayed how long the transmission effect maintain; Secondly, the coefficients
290 in VARMA model extended matrix measured the size of transmission effect between
291 cities.

292

293 **3 Results**

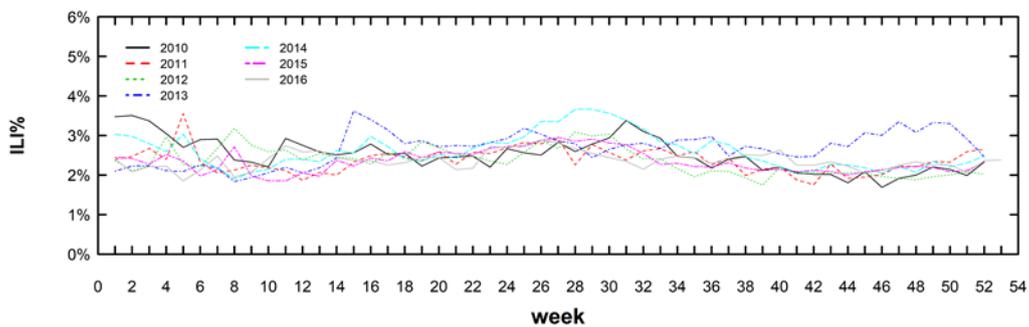
294 **3.1 The influenza situation in Sichuan between 2010-2016**

295 The total number of outpatients in Sichuan sentinel surveillance hospital from 2010 to
296 2016 was 31,898,487, and the total number of influenza-like cases in Sichuan was
297 784,984. As a result, the ILI% of Sichuan was 2.46% within six years. From 2010 to
298 2016, the cumulative ILI% was 2.51%, 2.36%, 2.38%, 2.69%, 2.60%, 2.34% and 2.36%.
299 The year with lowest ILI number was 2010 (84,766 visits), and highest was 2016
300 (137,945 visits); the lowest ILI% was in 2016 (2.36%) and the highest was in 2013
301 (2.69%). The 2010-2016 weekly ILI% distribution was shown in Fig.1 and Fig.2 below:



302

303 **Fig. 1** ILI% time series in Sichuan Surveillance Hospitals, 2010-2016



304

305 **Fig. 2** Yearly ILI% time series in Sichuan Surveillance Hospital, 2010-2016

306 From Fig. 1 and Fig. 2, the weekly ILI% in Sichuan province from 2010 to 2016 was
307 between 1.69% and 3.66%, which was consistent with past years. From Fig. 2, it could

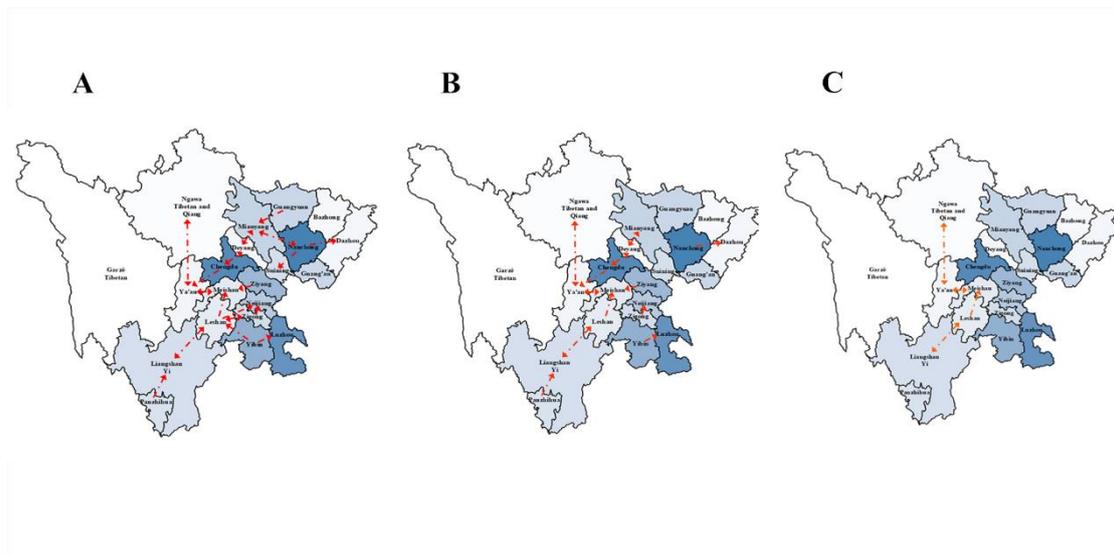
308 be seen there was a slight peak in winter and spring (From the 47th to 52nd week, and
309 from 1st to 9th week), similar to the result of the whole country.

310 3.2 Result of structure learning of possible influenza transmission route network

311 Combining the topological structure learning results in each unit (Additional file 4)

312 and order selection results in parameter learning part (Additional file 5), the final

313 influenza spatial transmission network was obtained as Fig 3:



314

315 **Fig.3** Sequential-week lagged transmission routes of influenza between different cities

316 by combining the results of structure learning and order selection results in each

317 learning unit. A)1-week lagged B) 2-week lagged C)3-week lagged

318 According to Twelfth Five-Year Plan on Urbanization Development of Sichuan

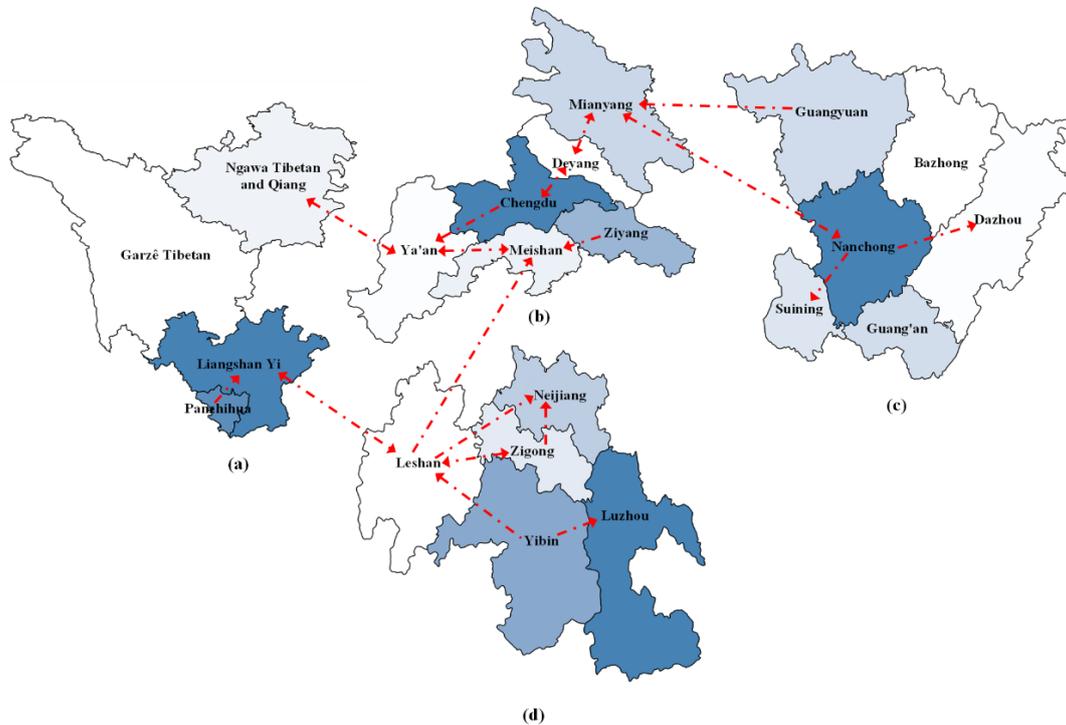
319 province, Sichuan province could be divided to 4 geographically meaningful

320 subregions: Chengdu Plain subregion (Chengdu, Mianyang, Deyang, Ziyang, Meishan

321 and Ya'an), Northeast Sichuan subregion (Bazhong, Guangyuan, Dazhou, Guang'an,

322 Suining and Nanchong), South Sichuan subregion(Neijiang, Zigong, Yibin, Leshan and

323 Luzhou), Pan West and other subregion(Panzhihua, Ganzi, Aba and Liangshan). We
 324 divided 1-week lagged transmission network of influenza to 4 subregions according to
 325 the criteria and showed this result in Fig.4.



326
 327 **Fig. 4** 1-week lagged influenza transmission network in different subregions of
 328 Sichuan province. (a)Pan west and other subregion (b)Chengdu Plain subregion (c)
 329 Northeast Sichuan subregion (d) South Sichuan subregion

330 From Fig.4, the cumulative ILI% in Luzhou city ranked first in the whole province,
 331 reminding us to focus on influenza management in Luzhou city. In general, if we
 332 defined the origin city of suspected influenza in Sichuan province as a city with single-
 333 direction arrows pointing to other cities in the province with no city in the province
 334 pointing to this city though. It could be seen from Fig.4 that in each subregion, a city
 335 of influenza origin could be detected, namely Ziyang in Chengdu Plain subregion,

336 Guangyuan in Northeast Sichuan, Yibin in South Sichuan, and Panzhihua in Pan West
337 and other subregion. These cities were all bordering cities in Sichuan province, playing
338 essential transportation hubs connecting to other provinces and could be guessed being
339 the birthplace of influenza in Sichuan province. Specifically, Ziyang was adjacent to
340 Chongqing municipality; Guangyuan bordered Gansu Province and Shaanxi Province;
341 Yibin City was a boundary of Sichuan, Yunnan and Guizhou Ministry; Panzhihua was
342 connected to Yunnan Province, suggesting there would be potential transmission routes
343 of influenza from bordering provinces or municipality.

344 From the results of structure learning in Additional file 4, there existed circular arrows
345 pointing to themselves universally in every city in Sichuan province, suggesting the
346 number of influenza-like cases in each city of Sichuan province had a temporal
347 autocorrelation, which was consistent with the results in most studies focusing on ILI%
348 time series using univariate time-series models such as the autoregressive moving
349 average model ARMA²³⁻²⁶, also consistent with the modeling ideas of Held, et al²⁶.

350

351 **3.3 Result of parameter learning of possible influenza transmission route network**

352 After managing parameter learning results in each unit, we summarized and displayed
353 possible transmission effects between different cities in each unit in Table 1, and the
354 specific parameter learning result using VARMA model was in Additional file 5.

355 **Table 1 summary of the parameters at sequential-lagged week between two adjacent cities in**

356 **each community**

Community		1-week lagged	2-week lagged	3-week lagged
No.	Spreading direction	effect	effect	effect
1	Chengdu to Deyang	-0.02	0.17	-
	Deyang to Chengdu	0.14	-0.05	-
2	Aba to Ya'an	0.13	-0.10	0.12
	Ya'an to Aba	0.06	-0.19	0.02
	Chengdu to Ya'an	-0.27	0.24	-0.37
	Ya'an to Aba	-0.67	0.13	-0.70
3	Deyang to Mianyang	0.00	-0.07	-
	Mianyang to Deyang	-0.14	0.26	-
4	Aba to Ya'an	0.02	-0.04	0.00
	Ya'an to Aba	0.06	-0.19	0.02
5	Nanchong to Dazhou	-0.11	-0.28	-
6	Chengdu to Deyang	-0.02	0.16	-
	Deyang to Chengdu	0.14	0.07	-
7	Chengdu to Ya'an	-0.38	0.34	-
	Ya'an to Meishan	-0.13	-0.03	-
8	Ziyang to Meishan	0.35	-0.09	-
9	Nanchong to Dazhou	-0.23	-0.23	-
10	Deyang to Mianyang	0.00	-0.08	-
	Mianyang to Deyang	-0.14	0.05	-
11	Mianyang to Nanchong	0.09	-0.09	-

Community		1-week lagged	2-week lagged	3-week lagged
No.	Spreading direction	effect	effect	effect
	Nanchong to Mianyang	0.00	0.16	-
12	Leshan to Liangshan	0.75	0.00	-0.90
	Leshan to Meishan	-0.83	-0.84	-0.16
	Liangshan to Leshan	0.04	-0.05	-0.01
	Meishan to Ya'an	0.07	-0.02	-0.03
	Ya'an on Meishan	0.53	-0.38	0.11
13	Leshan to Liangshan	0.33	-0.43	-
	Liangshan to Leshan	0.04	-0.05	-
	Yibin to Leshan	-0.01	-0.03	-
14	Leshan to Meishan	-0.05	-	-
	Leshan to Neijiang	-0.10	-	-
	Leshan to Zigong	-0.09	-	-
	Zigong to Leshan	-0.02	-	-
15	Leshan to Zigong	-0.10	-	-
	Zigong to Leshan	-0.13	-	-
16	Panzhuhua to Liangshan	0.17	-	-
17	Zigong to Neijiang	-0.02	-0.31	-
18	Yibin to Luzhou	-0.07	-0.04	-
19	Ziyang to Meishan	0.00	-0.27	-

Community No.	Spreading direction	1-week lagged effect	2-week lagged effect	3-week lagged effect
20	Mianyang to Nanchong	0.07	-	-
	Nanchong to Mianyang	0.15	-	-
	Nanchong to Suining	-0.10	-	-

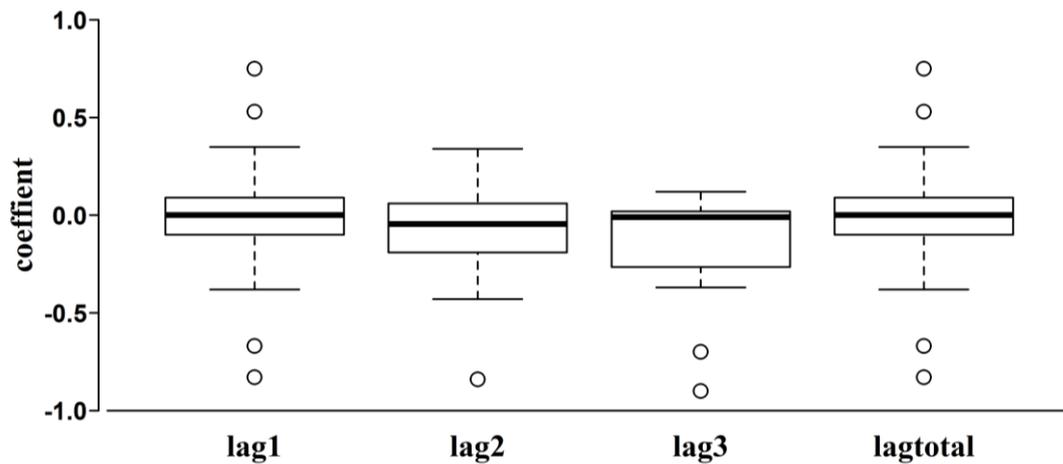
357

358 After conducting descriptive statistical analysis, we displayed the result at sequential-
359 lagged week in Table 2.

360 **Table 2 Statistical description of parameters at sequential-lagged week**

Lagged week	Number of Independent variables	Min	Lower Quartile	Median	Upper Quartile	Maximum
1 week	42	-0.8300	-0.1025	0.0000	0.0900	0.7500
2 week	32	-0.8400	-0.1900	-0.0450	0.0650	0.3400
3 week	11	-0.9000	-0.3700	-0.0100	0.0200	0.1200
Total	85	-0.9000	-0.1300	-0.0200	0.0700	0.7500

361 The corresponding box diagram was as follows:



362

363 **Fig. 5** Box diagram of the statistical description of parameters at sequential-lagged week

364

between two adjacent cities in each community

365 **4 Discussion**

366 According to the complexity of the structure learning network (the number of direct
 367 edge arrows) in each learning unit, influenza transmission network was most developed
 368 in Chengdu plain and Southern Sichuan, followed by northeastern Sichuan, and Pan
 369 West and other regions. Considering the fact influenza was a typical human-to-human
 370 infectious disease, it is reasonable to associate the extent of its transportation network
 371 coverage with the complexity of the structural structure learning network. According to
 372 *Sichuan Transportation Yearbook 2017*, Sichuan's highway passenger turnover was
 373 59.78 billion person*kilometres, ranking first in all traffic methods in 2016, meaning
 374 the transportation between cities within Sichuan province was dominated by highway
 375 transportation. In terms of the result, we could find that except for the complexity of
 376 the influenza transmission network in the northeast Sichuan, which was not consistent
 377 with the coverage of the trunk road network, in the remaining areas, the influenza

378 transmission network complexity was consistent with the coverage of the highway
379 network. Specifically, the number of suspected influenza transmission routes in
380 Chengdu Plain, Northeast Sichuan, South Sichuan and Pan West and other regions were
381 10, 4, 8 and 3; the total passenger number was 349,340 279,490 342,710 and 184,650
382 thousand, respectively; the number of passenger turnover was 2,246,472, 13,915,980,
383 15,535,290, and 7868,210 thousand person-kilometers. It showed that the complexity
384 of the network was highly correlated with the number of passengers and the number of
385 passengers.

386 Furthermore, after checking highway routes within Sichuan province probably playing
387 essential role in the transmission of influenza from Baidu map, we found possible
388 essential routes as below:

389 In Chengdu Plain subregion, the Ziyang→Meishan route might originate from the
390 Suizimei Expressway, and the Chengdu→Ya'an propagation route was to some extent
391 in accord with the Chengya Expressway Direction; while the
392 Guangyuan→Mianyang←→Deyang←→Chengdu→Ya'an propagation route was
393 possibly in line with G5 Jingkun Expressway (the Mianguang section, the Chengmian
394 section and the Chengya section), and the Nanchong ←→ Mianyang ←→ Deyang
395 ←→ Chengdu route was expected to be brought by the Chengdemiannan Expressway.
396 As another source, Leshan→Meishan←→Ya'an propagation route was guessed to be
397 caused by Leya Expressway, which was another important highway flowing through
398 Ya'an.

399 In Northeast Sichuan subregion, as could be seen from the map, Guangyuan was
400 adjacent to both Shaanxi province and Gansu province, and National Highway 108
401 played an important channel for Shaanxi, Ningxia and Gansu entering Sichuan province.
402 After the possible spreading of influenza from Guangyuan to Mianyang, another route
403 from Mianyang to Nanchong was suspected to be spread via the Chengdemiannan
404 Expressway. It was speculated that the propagation route from Nanchong to Suining
405 and Dazhou was caused by National Highway 318 and the Suining-Nanchong and Suixi
406 sections of G42 Rong Expressway.

407 In South Sichuan subregion, the main transmission routes were Yibin→Luzhou,
408 Yibin→Leshan→Neijiang, and Yibin→Leshan←→Zigong→Neijiang. As was
409 commonly known, Yibin played an essential transportation hub in southern Sichuan
410 transportation network, and this city was a junction of Sichuan, Yunnan and Guizhou
411 provinces with S206 provincial road crossing it. Therefore, it could be guessed that
412 Yibin was one of the origins of influenza-like cases in Sichuan province, and the source
413 was mainly from Yunnan. The main highways include Neiyi Expressway, Leyi
414 Expressway and Yilu Expressway might explain why there were propagation routes
415 between Yibin, Leshan and Luzhou. The propagation routes between Leshan and
416 Meishan was guessed by Chengle Expressway and Lemei Expressway; the route
417 between Leshan and Yibin might be brought by Leyi Expressway; the propagation route
418 between Leshan and Zigong might be brought by Lezi Expressway. The propagation
419 route between Zigong and Neijiang should be formed by Neizi Expressway, Chengzilu
420 and Zilong highway.

421 In Pan West and other subregion, there existed G5 Beijing-Kunming Expressway across
422 Panzhihua as a junction between Sichuan and Yunnan, and this expressway could be
423 regarded as an important entrance from Yunnan to Sichuan. In terms of the spreading
424 route from Panzhihua to Liangshan, G5 Beijing-Kunming Expressway was guessed to
425 play an essential role.

426 To summarize, the spreading routes learning from DBN network and VARMA model
427 were almost in accordance with the traffic network of Sichuan province, providing a
428 scientific proof supporting the rationality of influenza transmission network of Sichuan.
429 On other hand, influenza was a typical human-to-human infectious disease, so logically
430 transportation lines should play an important role in the transmission of influenza. The
431 important role of transportation has been confirmed by a large number of studies, for
432 example, Hidenori²⁷ found that traffic control can delay the spread of flu during peak
433 flu periods in a simulation study.

434 According to the condition of using Dynamic Bayesian network, the predictions of
435 fitting DBN and VARMA models must be: 1) the infection process should be first-order
436 Markovian; 2) the case number of time t must depend on time $t-1$ and not depend on
437 vectors of time t ; 3) the time-lagged association of vectors must be stationary. In one
438 word, the method of DBN and VARMA suits the construction of transmission network
439 of long-period stably spreading infectious disease without too much prevention
440 intervention.

441 In the perspective of prevention for influenza, from Fig. 3, during 3-week influenza

442 transmission cycle, the number of suspected influenza transmission routes in Sichuan
443 province showed a decreasing trend. From Fig. 3, it could be speculated that the period
444 of influenza transmission cycle was longer in western Sichuan and Chengdu plain and
445 shorter in the northeastern Sichuan, suggesting that we should consider appropriately
446 extending the period for influenza prevention and control in the western Sichuan and
447 Chengdu plains compared with the northeastern Sichuan. Furthermore, from Fig. 5, in
448 terms of the interquartile range at different lagged time, it could be seen that most of
449 the parameters were in the range of $-0.13\sim 0.07$, indicating that the suspected influenza
450 transmission relationship between cities in each community was in general not obvious.
451 However, there were several routes between cities worthy of attention: the spreading
452 relationship between Leshan and Liangshan was 0.75 after 1 week, the relationship
453 between Ya'an and Meishan after 1 week was 0.53, the relationship between Leshan
454 and Meishan after 1 week was -0.83 , with the relationship after 2 weeks was -0.84 , and
455 the transmission relationship between Ya'an and Aba after 1 week was -0.38 with the
456 relationship after 3 weeks was -0.70 . The result indicated the estimated time-lagged
457 relationship of influenza cases between these cities were relatively close, suggesting
458 that prevention of influenza transmission between these cities should be paid attention
459 to. In addition, it could be seen from a lag of 1 week to a lag of 3 weeks, the proportion
460 of negative parameters was gradually increasing, with the degree of dispersion was
461 gradually increasing, which was guessed to relate to herd immunity. On other side,
462 because the disease resistance of populations in each city was different, parameters
463 representing the transmission relationship between cities over time varied, leading to

464 the increase of the degree of dispersion.

465

466 **5 Conclusions**

467 This study proposed an innovative framework for exploring the potentially stable
468 transmission routes between different regions and measuring specific size of the
469 transmission effect. In comparison with a series of studies using constructed complex
470 network combining with key transmission dynamics parameters predicting the
471 incidence of one infectious disease in one area, our study explored this complex
472 network using statistically framework with real-world surveillance data themselves
473 referencing to the framework of constructing gene regulatory network. The result could
474 be used for the infectious disease key area confirmation by considering their adjacent
475 areas' incidence and the transmission relationship. To be specific, the potential 1-week
476 lagged transmission effect from Leshan to Liangshan was 0.75, but from Leshan to
477 Meishan was -0.05, implying that once an influenza epidemic happened in Leshan, the
478 resource allocation could be considered for Liangshan Yi primarily rather than Meishan.

479 The framework of this study suits the transmission network exploration of long-period
480 stably infectious disease without too much prevention intervention considering the
481 precondition of unchanged dynamic dependencies between vectors when fitting
482 VARMA and DBN model. Some limitation must be implicated: Firstly, we did not
483 consider the multiple test correction so Type I error increases combining networks from
484 different learning units together. Second, the assumption that non-adjacent cities'

485 transmission effect could be neglected is not strong, and more network construction
486 methods including network clustering, prediction modelling based on network. Besides,
487 we used the estimated ILI case numbers rather than influenza confirmed case numbers
488 for fitting those models, leading to exaggerated relationship estimation of influenza
489 transmission.

490

491 **Declaration:**

492 *Ethics approval and consent to participate: not applicable*

493 *Consent for publication: not applicable*

494 *Availability of data and materials: The data that support the findings of this study are*
495 *available from Sichuan Center for Disease Control and Prevention but restrictions*
496 *apply to the availability of these data, which were used under license for the current*
497 *study, and so are not publicly available. Data are however available from the authors*
498 *upon reasonable request and with permission of Sichuan Center for Disease Control*
499 *and Prevention.*

500 *Competing interests: The authors declare that they have no competing interests*

501 *Funding: Tao Zhang's work was supported by Sichuan Science and Technology*
502 *Program (2020YFS0015, 2020YFS0091).*

503 *Authors' contributions: JQQ analyzed the data and was a major contributor in writing*
504 *the manuscript. All authors read and approved the final manuscript, HMW helped*

505 *perform the analysis with constructive discussions, TZ contributed significantly to*
506 *analysis and manuscript preparation, and CHY collected and provided us with the data.*

507 **Acknowledgements:** *Many thanks are expressed to deceased professor*

508 *Xiaosong Li in Sichuan University for his kind guide in the manuscript, and to Ting*

509 *Huang in Sichuan center for disease control and prevention for her some ideas.*

510

511

512

513

514 **References:**

515 1. Su, S., Fu, X., Li, G., Kerlin, F. & Veit, M. Novel Influenza D virus: Epidemiology,
516 pathology, evolution and biological characteristics. *Virulence* **8**, 1580–1591 (2017).

517 2. Influenza (Seasonal). [https://www.who.int/news-room/fact-sheets/detail/influenza-](https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal))
518 (seasonal).

519 3. Sah, P. *et al.* Future epidemiological and economic impacts of universal influenza
520 vaccines. *Proc Natl Acad Sci USA* **116**, 20786–20792 (2019).

521 4. Grohskopf, L. A. *et al.* Prevention and Control of Seasonal Influenza with Vaccines:
522 Recommendations of the Advisory Committee on Immunization Practices - United
523 States, 2019-20 Influenza Season. *MMWR Recomm Rep* **68**, 1–21 (2019).

524 5. Petrie, J. G. *et al.* Modest Waning of Influenza Vaccine Efficacy and Antibody Titers

- 525 During the 2007–2008 Influenza Season. *J Infect Dis.* **214**, 1142–1149 (2016).
- 526 6. Hill, E. M., Petrou, S., de Lusignan, S., Yonova, I. & Keeling, M. J. Seasonal
527 influenza: Modelling approaches to capture immunity propagation. *PLoS Comput*
528 *Biol* **15**, e1007096 (2019).
- 529 7. Fu, X., Small, M. & Chen, G. *Propagation Dynamics on Complex Networks.* (2014).
- 530 8. Shi, B. *et al.* Identifying key bird species and geographical hotspots of avian
531 influenza A (H7N9) virus in China. *Infect Dis Poverty* **7**, 97 (2018).
- 532 9. WHO | WHO surveillance case definitions for ILI and SARI. *WHO*
533 [http://www.who.int/influenza/surveillance_monitoring/ili_sari_surveillance_case_d](http://www.who.int/influenza/surveillance_monitoring/ili_sari_surveillance_case_definition/en/)
534 [efinition/en/](http://www.who.int/influenza/surveillance_monitoring/ili_sari_surveillance_case_definition/en/).
- 535 10. Alonso, W. J. *et al.* Seasonality of influenza in Brazil: a traveling wave from the
536 Amazon to the subtropics. *American Journal of Epidemiology* **165**, 1434–1442
537 (2007).
- 538 11. Paul, M., Held, L. & Toschke, A. M. Multivariate modelling of infectious disease
539 surveillance data. *Statist. Med.* **27**, 6250–6267 (2008).
- 540 12. Nagarajan, R., Scutari, M. & Lèbre, S. *Bayesian Networks in R.* (Springer New
541 York, 2013). doi:10.1007/978-1-4614-6446-4.
- 542 13. Sandri, M., Berchialla, P., Baldi, I., Gregori, D. & De Blasi, R. A. Dynamic
543 Bayesian Networks to predict sequences of organ failures in patients admitted to ICU.
544 *J Biomed Inform* **48**, 106–113 (2014).
- 545 14. Kourou, K., Rigas, G., Exarchos, K. P., Papaloukas, C. & Fotiadis, D. I. Prediction
546 of oral cancer recurrence using dynamic Bayesian networks. in *Engineering in*

- 547 *Medicine & Biology Society* (2016).
- 548 15. Zhang, T. *et al.* Dynamic Bayesian network in infectious diseases surveillance: a
549 simulation study. *Sci Rep* **9**, 10376 (2019).
- 550 16. Bueno, M. L., Hommersom, A., Lucas, P. J., Lappenschaar, M. & Janzing, J. G.
551 Understanding disease processes by partitioned dynamic Bayesian networks. *J*
552 *Biomed Inform* **61**, 283–97 (2016).
- 553 17. Haddawy, P. *et al.* Spatiotemporal Bayesian networks for malaria prediction. *Artif*
554 *Intell Med* **84**, 127–138 (2018).
- 555 18. TIBSHIRANI, R. Regression Shrinkage and Selection via the Lasso. 23.
- 556 19. Opgen-Rhein, R. & Strimmer, K. Learning causal networks from systems biology
557 time course data: an effective model selection procedure for the vector
558 autoregressive process. *BMC Bioinformatics* **8**, (2007).
- 559 20. Lèbre, S. Inferring Dynamic Genetic Networks with Low Order Independencies.
560 *Statistical Applications in Genetics and Molecular Biology* **8**, 1–38 (2009).
- 561 21. Chiquet, J., Smith, A., Grasseau, G., Matias, C. & Ambroise, C. SIMoNe:
562 Statistical Inference for MODular NETworks. *Bioinformatics* **25**, 417–418 (2009).
- 563 22. Tsay, R. S. Multivariate Time Series Analysis. 522.
- 564 23. Chadsuthi, S., Iamsirithaworn, S., Triampo, W. & Modchang, C. Modeling
565 Seasonal Influenza Transmission and Its Association with Climate Factors in
566 Thailand Using Time-Series and ARIMAX Analyses. *Computational and*
567 *mathematical methods in medicine* **2015**, 436495–436495 (2015).
- 568 24. Kane, M. J., Price, N., Scotch, M. & Rabinowitz, P. Comparison of ARIMA and

- 569 Random Forest time series models for prediction of avian influenza H5N1 outbreaks.
570 *BMC bioinformatics* **15**, 276–276 (2014).
- 571 25. Song, X. *et al.* Time series analysis of influenza incidence in Chinese provinces
572 from 2004 to 2011. *Medicine* **95**, e3929–e3929 (2016).
- 573 26. Held, L., Hofmann, M., Höhle, M. & Schmid, V. A two-component model for
574 counts of infectious diseases. *Biostatistics* **7**, 422 (2006).
- 575 27. Yasuda, H. *et al.* Preparedness for the Spread of Influenza: Prohibition of Traffic,
576 School Closure, and Vaccination of Children in the Commuter Towns of Tokyo.
577 *Journal of Urban Health* **85**, 619–635 (2008).
- 578