Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Global influenza surveillance with Laplacian multidimensional scaling*

Xi-chuan ZHOU†[1,2], Fang TANG[1,3], Qin LI[4], Sheng-dong HU[1,3],
Guo-jun LI[5], Yun-jian JIA[1], Xin-ke LI[1], Yu-jie FENG[6]

([1]*College of Communications Engineering, Chongqing University, Chongqing 400044, China*)
([2]*MOE Key Laboratory of Dependable Service Computing in Cyber Physical Society, Chongqing 400044, China*)
([3]*Chongqing Engineering Laboratory of High Performance Integrated Circuits, Chongqing University, Chongqing 400044, China*)
([4]*Centers for Disease Control and Prevention of Chongqing, Chongqing 400042, China*)
([5]*Chongqing Communication Institute, Chongqing 400030, China*)
([6]*Third Military Medical University, Southwest Hospital, Chongqing 400030, China*)
†E-mail: zxc@cqu.edu.cn

Received Oct. 16, 2015; Revision accepted Feb. 17, 2016; Crosschecked Apr. 28, 2016

Abstract: The Global Influenza Surveillance Network is crucial for monitoring epidemic risk in participating countries. However, at present, the network has notable gaps in the developing world, principally in Africa and Asia where laboratory capabilities are limited. Moreover, for the last few years, various influenza viruses have been continuously emerging in the resource-limited countries, making these surveillance gaps a more imminent challenge. We present a spatial-transmission model to estimate epidemic risks in the countries where only partial or even no surveillance data are available. Motivated by the observation that countries in the same influenza transmission zone divided by the World Health Organization had similar transmission patterns, we propose to estimate the influenza epidemic risk of an unmonitored country by incorporating the surveillance data reported by countries of the same transmission zone. Experiments show that the risk estimates are highly correlated with the actual influenza morbidity trends for African and Asian countries. The proposed method may provide the much-needed capability to detect, assess, and notify potential influenza epidemics to the developing world.

Key words: Surveillance gap, Influenza, Spatial-transmission model
http://dx.doi.org/10.1631/FITEE.1500356 **CLC number:** TP391.4

1 Introduction

The Global Influenza Surveillance Network (GISN) of the World Health Organization (WHO) is an essential foundation for monitoring and managing an influenza pandemic. At present, the GISN is collecting and processing influenza virological data at the global scale, which comprises 136 national influenza centers in 106 countries and 5 collaborating centers for reference and research on influenza (http://www.influenzacentre.org/centre_GISN.htm). The GISN system has proven to be valuable, but it leaves severe spatial gap, principally in Asia and Africa (Fig. 1). Moreover, only 21 countries with national surveillance networks are reporting their influenza surveillance data to the WHO in a standard form (http://www.who.int/influenza/

* Project supported by the National Natural Science Foundation of China (Nos. 61103212 and 61471073) and the Chinese Post-Doctoral Science Foundation (Nos. 2012M521678 and 2013T60836)
ORCID: Xi-chuan ZHOU, http://orcid.org/0000-0002-3304-3045

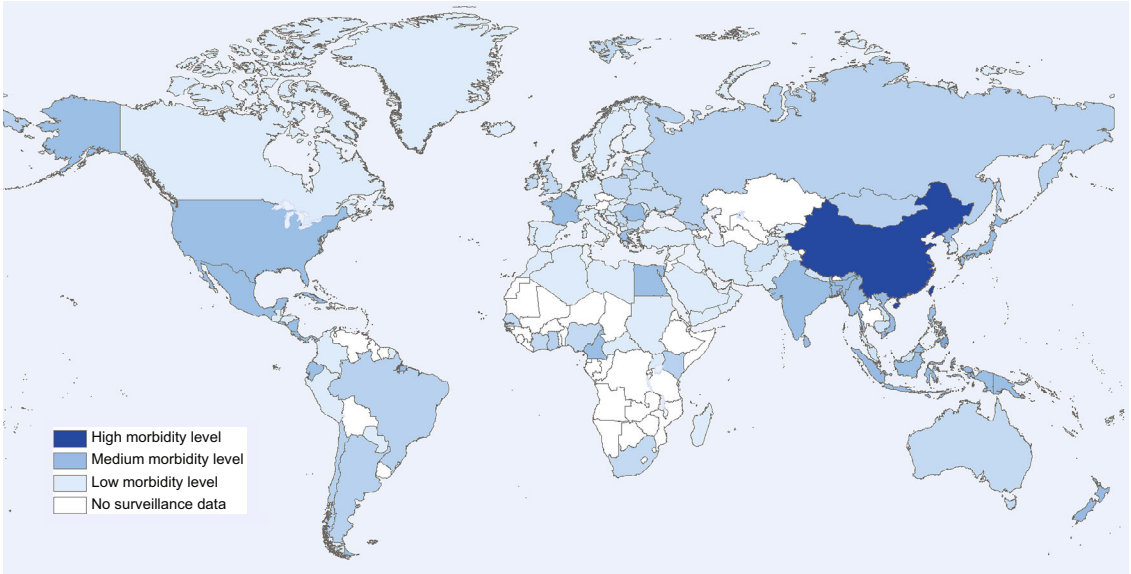

Fig. 1 Global influenza morbidity levels reported by the WHO in Feb. 2010

gisrs_laboratory/flunet/en/). For most developing countries, the data are available only discontinuously.

Our work attempts to map the influenza epidemic risk of different countries at the global scale. There are a variety of reasons for mapping the geographical distribution of the influenza epidemic risk (Lavanchy, 1999). First, it allows an immediate visualization of the extent and magnitude of the public health problem (He *et al.*, 2015b). Second, when based on empirical evidence, maps could support carefully weighted assessments by decision makers on the advantages and disadvantages of alternative courses of action (Hay *et al.*, 2013; Nelson *et al.*, 2015). These may range from helping plan international intervention strategies (Kenah *et al.*, 2011) to advice for individuals on whether to vaccinate and provide prophylaxis before travel (http://www.who.int/ith/other_health_risks/infectious_diseases/en/).

The 2009 influenza A(H1N1) pandemic taught us that monitoring influenza epidemic risk at the global scale is of great importance; however, little research focused on resource-limited countries due to the lack of surveillance data (Briand *et al.*, 2014). Since 2009, there has been a rapid growth of interest in developing large-scale spatial-transmission models for influenza surveillance (He *et al.*, 2013). There were roughly four groups of spatial-transmission models categorized by different assumptions about the characteristic of disease transmission between different members (Riley, 2007). Each member could represent an individual, a household, or a country, depending on the scale of research. For the multi-group models, the disease could transmit only between members in the same group (Ferguson *et al.*, 2005; Longini *et al.*, 2005). For network models, the infection risk of a member is zero unless it connects to an infected member in the transmission network (Eubank *et al.*, 2004). The distance-based models assume that any infected member could infect all susceptible members within range (Ferguson *et al.*, 2001; Keeling *et al.*, 2001). The pairwise probability of infection is usually a monotonically decreasing function of distance. The patch models are similar to multi-group models, except that the members in the home-patch could also be infected by members in adjacent or close patches (Cooper *et al.*, 2006; Hollingsworth *et al.*, 2007). Wang and Li (2014) introduced the metapopulation data and summarized relevant progress in recent years. Besides the spatial-transmission models, traditional Bayesian and Markov models were used extensively in epidemic risk estimations (Best *et al.*, 2005; Zhou *et al.*, 2013). These models have been successfully used in visualizing and mapping epidemic risks at different scales. However, they cannot be applied directly to regions where no surveillance data are available.

At the global scale, the epidemic risk of the

influenza varies significantly from continent to continent. For example, most H5N1 and H7N9 cases were reported in Asia. Research indicated that the level of economic development also plays a part in influenza transmission, which partly explains why the developing countries in Asia and Africa had higher risks of influenza epidemics (Oshitani *et al.*, 2008). On the other hand, despite the fact that influenza risks vary from country to country, recent research indicated that the influenza dynamics of many countries are strongly correlated in time after the 2009 H1N1 pandemic (He *et al.*, 2015b). Similar climate or environment might result in similar patterns of influenza transmission in nearby countries (Tamerius *et al.*, 2013). At present, the WHO divides the world into 18 influenza transmission zones according to similar influenza transmission patterns (WHO, 2014). Intuitively, if a country lacks surveillance data, one might be able to estimate its epidemic risk using the surveillance data for countries in the same transmission zone. As a special case, He *et al.* (2015a) approximated the influenza morbidity in Saudi Arabia, which was unavailable, using the data collected from nearby countries.

This paper presents a novel spatial-transmission model to estimate the influenza epidemic risk which could be measured by different means, such as the morbidity or the reproductive number of the influenza. We use the simple measure of the total morbidity to represent the influenza epidemic risk, which has been widely used in previous research (Zhou and Shen, 2010; Zhou *et al.*, 2011; He *et al.*, 2015b). To extend our model to countries where no surveillance data are available, we build a hybrid spatial-transmission model according to the characteristics of influenza transmission. At the global scale, a network model is first built over countries where surveillance data are available. We use the classic multi-dimensional scaling (MDS) formulation and attempt to preserve the spatial difference of influenza morbidity between different countries. Then, we add a second layer of the multi-group model over all countries including those with no surveillance data. We define 18 groups, each group representing an influenza transmission zone. Assuming that the countries in the same transmission zone have similar epidemic risks, we minimize the differences of epidemic risks among countries in the same transmission zone. Finally, we integrate the network model and the multi-group model into one optimization formulation and transform it into a generalized eigenvalue problem.

2 Data and material

The WHO gathers the influenza morbidity data of 130 countries (or regions) and has published the data by the FluNet (http://www.who.int/influenza/gisrs_laboratory/flunet/en/). The morbidity is of weekly resolution and is divided by subtypes, i.e., subtype A(H5), A(H1), A(H1N1)pdm09, A(H3), A(not subtyped), B(Yamagata lineage), B(Victoria lineage), and B(Lineage not determined). A typical example of the morbidity examined is shown in Fig. 2. Our model is built and evaluated using the WHO influenza morbidity in a week-by-week fashion, and the period covered by our experiments is 260 weeks from Jan. 2009 to Dec. 2013.

At present, the GISN has covered most countries in the western world. However, the influenza surveillance data for 28 (of 56) African and 15 (of 48) Asian countries (or regions) remain unavailable. Moreover, of all the countries and regions reporting surveillance data to the WHO, only 21 countries have their own surveillance networks, allowing a standardized form of reporting. However, for most countries, the surveillance data are incomplete. For the countries covered by the GISN, the average missing rates are 26.50% for African countries and 24.22% for Asian countries.

3 Method

Our method is designed to estimate the epidemic risks by preserving both global variations and regional similarities in disease transmission. Suppose $x_{i,p}$ is the morbidity number of influenza subtype p ($p \leq 8$) for the ith country (region). The spatial difference of influenza morbidity between countries i and j is defined using the usual squared difference as

$$\delta_{i,j} = \sum_{p=1}^{8} (x_{i,p} - x_{j,p})^2. \tag{1}$$

Suppose y_i indicates the epidemic risk of country i, and k is the total number of countries and regions where influenza surveillance data are available. To estimate y_i and preserve the spatial variations of epidemic risk, we employ the MDS formulation as

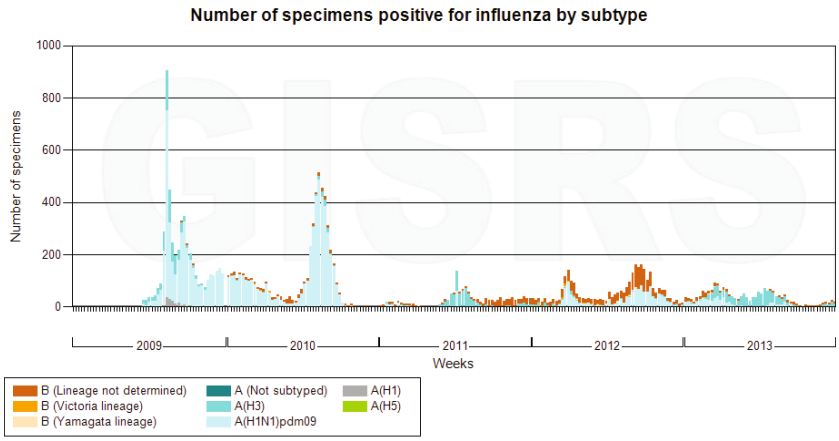

Fig. 2 Typical example of the influenza morbidity of India published by the FluNet (This image was used here with WHO's copyright permission. References to color refer to the online version of this figure)

(Williams, 2002)

$$\min_{y_1,y_2,\ldots,y_k} \sum_{i=1}^{k}\sum_{j=1}^{k}(\|y_i-y_j\|^2-\delta_{i,j})^2, \tag{2}$$

to calculate the epidemic risks such that the morbidity difference between countries is preserved as well as possible.

To combine the above optimization with the regional similarity criterion, we transform problem (2) into its dual maximization form. Suppose $\boldsymbol{\Delta}=[\delta_{i,j}]_{k\times k}$ is the risk variation matrix, $\boldsymbol{e}$ the column vector of ones, $\boldsymbol{I}$ the identity matrix, and $\boldsymbol{K}$ the standardized dissimilarity matrix written as

$$\boldsymbol{K}=-\frac{1}{2}(\boldsymbol{I}-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e})\boldsymbol{\Delta}(\boldsymbol{I}-\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e}).$$

It has been shown in Williams (2002) that, given the epidemic risk vector $\boldsymbol{y}=[y_1,y_2,\ldots,y_k]^{\mathrm{T}}$, the solution of problem (2) is $\boldsymbol{y}=\sqrt{\alpha}\boldsymbol{w}$ where α and $\boldsymbol{w}$ are the first eigenvalue and eigenvector of $\boldsymbol{K}$, respectively. Suppose the length of $\boldsymbol{y}$ is restricted, e.g., constrained by a constant as $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}=\alpha$. Problem (2) can be written equivalently as the following maximization form:

$$\max_{\boldsymbol{y}} \boldsymbol{y}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{y}. \tag{3}$$

Suppose only k out of n countries (or regions) report their influenza surveillance data to the WHO. Given the extended matrix $\overline{\boldsymbol{K}}$ defined as

$$\overline{\boldsymbol{K}}=\begin{pmatrix}\boldsymbol{K} & \boldsymbol{0}_{k,n-k}\\ \boldsymbol{0}_{n-k,k} & \boldsymbol{0}_{n-k,n-k}\end{pmatrix}_{n\times n},$$

problem (3) can be written equally as

$$\max_{\overline{\boldsymbol{y}}} \overline{\boldsymbol{y}}^{\mathrm{T}}\overline{\boldsymbol{K}}\overline{\boldsymbol{y}}. \tag{4}$$

Note that $\overline{\boldsymbol{y}}=[\boldsymbol{y}^{\mathrm{T}},y_{k+1},\ldots,y_n]^{\mathrm{T}}$ includes the epidemic risks for $n-k$ countries (or regions) where no surveillance data are available for the examined week. To estimate $\overline{\boldsymbol{y}}$, we assume that the countries in a same influenza transmission zone have similar epidemic risks. Suppose $\boldsymbol{S}=[s_{ij}]_{n\times n}$ represents the transmission similarity matrix and

$$s_{ij}=\begin{cases}0, & \mathcal{Z}_i\neq\mathcal{Z}_j, \quad (5a)\\ 1, & \mathcal{Z}_i=\mathcal{Z}_j, \quad (5b)\end{cases}$$

where $\mathcal{Z}_i\in\{1,2,\ldots,18\}$ indicates which influenza transmission zone the ith country is in. To preserve the geographical similarity of influenza transmission in each zone, we minimize the squared difference of epidemic risks for countries in the same zone as

$$\min_{y_1,y_2,\ldots,y_n} \sum_{i,j=1}^{n} s_{ij}(y_i-y_j)^2. \tag{6}$$

Suppose $\boldsymbol{D}=[d_{i,j}]_{n\times n}$ is a diagonal matrix where $d_{ii}=\sum_{j=1}^{n}s_{ij}$. The matrix $\boldsymbol{L}=\boldsymbol{I}-\boldsymbol{D}$ is known as the Laplacian matrix. By simple reduction, problem (6) can be written as

$$\min_{\overline{\boldsymbol{y}}} \overline{\boldsymbol{y}}^{\mathrm{T}}\boldsymbol{L}\overline{\boldsymbol{y}}. \tag{7}$$

Since our method aims to preserve both global variations and regional similarities of the influenza epidemic risk, we attempt to maximize problem (4)

and minimize problem (7) simultaneously, which can be achieved by

$$\max_{\overline{\boldsymbol{y}}} \frac{\overline{\boldsymbol{y}}^{\mathrm{T}} \overline{\boldsymbol{K}} \overline{\boldsymbol{y}}}{\overline{\boldsymbol{y}}^{\mathrm{T}} \boldsymbol{L} \overline{\boldsymbol{y}}}. \tag{8}$$

The objective function (8), known as the generalized Rayleigh quotient, is equivalent to the following generalized eigenvalue problem:

$$\overline{\boldsymbol{K}} \overline{\boldsymbol{y}} = \lambda \boldsymbol{L} \overline{\boldsymbol{y}}. \tag{9}$$

The maximum of function (8) is achieved when $\overline{\boldsymbol{y}}$ is the eigenvector corresponding to the largest eigenvalue of the above generalized eigenvalue problem. It is worth noting that, since we use the largest general eigenvalue as the estimated risk, matrix $\boldsymbol{L}$ does not have to be of full rank for computation.

4 Experiments and results

One important hypothesis of our method is the high correlation of epidemic risks among countries in the same influenza transmission zone. There are several widely used measures to evaluate the epidemic risk, such as the total morbidity, basic reproductive number, and effective reproductive number. Guided by Occam's Razor, we use the simple measure of the total morbidity to represent the actual influenza epidemic risk, which has been widely used in previous research (Best *et al.*, 2005; Hay *et al.*, 2013).

To examine the hypothesis, we arrange the actual influenza epidemic risk for each country through the examined period as a vector, and calculate the Pearson correlation coefficients of the epidemic risks between all 130 countries (regions) examined. The average correlation coefficients between countries in the same transmission zone range from 0.75 (Western Africa zone, $\sigma = 0.05$) to 0.90 (North Africa zone, $\sigma = 0.05$) (except the Southern Africa zone which has only one country), which are significantly higher than the average correlation of all countries examined ($\rho = 0.15, \sigma = 0.05$).

To examine the Laplacian MDS method, we compare the estimated risk with the actual risk measurement for all the countries covered by the GISN. Fig. 3a shows an example of the estimated risk for different countries in the 8th week of 2010. The top 10 countries with the highest estimated risks were China, Japan, Egypt, Romania, USA, Russian, Mexico, Greece, Singapore, and France, which were consistent with the actual risk measurements.

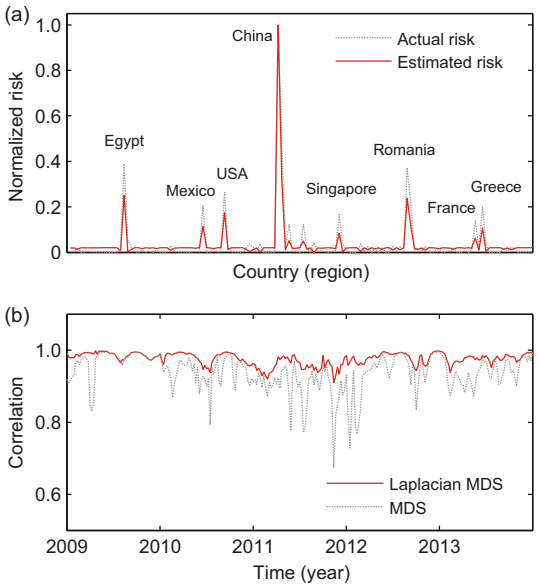

Fig. 3 Results for the countries covered by the GISN: (a) comparison of the estimated risk and the actual risk measurement in the 8th week of 2010; (b) the correlation between the estimated risks and the actual risk measurements of different countries for all the examined weeks from Jan. 2009 to Dec. 2013

The full correlation results are shown in Fig. 3b, in which we compare the Laplacian MDS (solid line, $0.90 \leq \rho \leq 0.99$, $\sigma = 0.05$) with the classic MDS (dotted line, $0.69 \leq \rho \leq 0.98$, $\sigma = 0.05$). Generally, the average correlation between the estimated risk and the actual risk measurement is strong for both the Laplacian MDS ($\rho = 0.93$, $\sigma = 0.05$) and the classic MDS ($\rho = 0.89$, $\sigma = 0.05$). The performance of the Laplacian MDS seems more stable than that of the classic MDS. The reason might be that the output of the classic MDS does not have to be positive, which could result in an underestimation of the epidemic risk. On the other hand, since the Laplacian MDS attempts to smooth the output, it leads to positive estimates and more stable results for epidemic risk mapping.

By taking advantage of the correlations between countries in the same transmission zone, the Laplacian MDS can be used to fill the surveillance gaps in African and Asian countries (areas). Fig. 4 shows an example of the global influenza epidemic risk estimation. The risk estimates for African countries during the examined period indicate that, though no cases were reported, the influenza epidemic risks in unmonitored African countries during the H1N1 pandemic period are not negligible, which is consistent with later findings (WHO Regional Office for Africa,

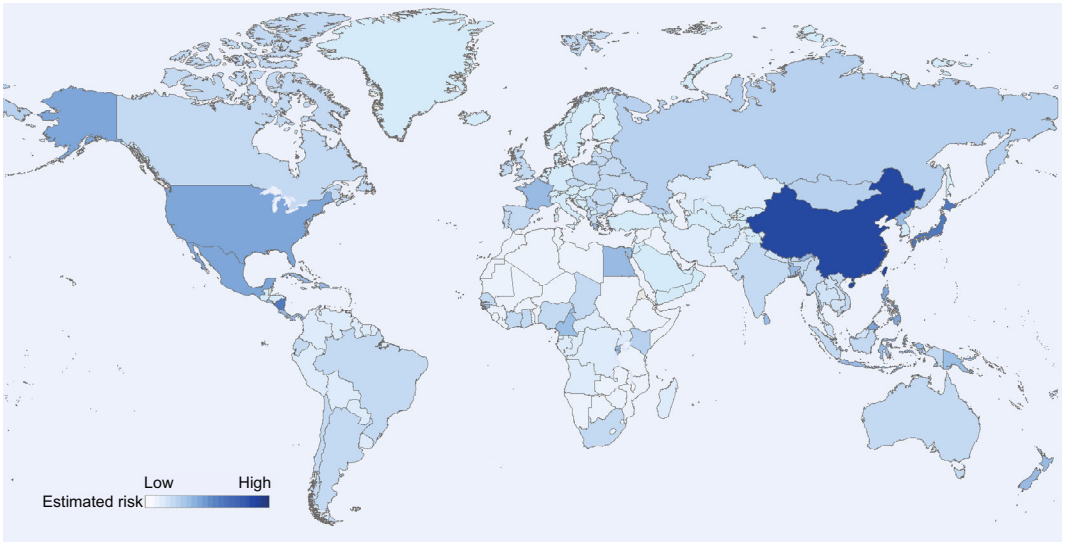

Fig. 4 The estimated influenza epidemic risks in Feb. 2010

2009). Thanks to the high correlation between the estimated risk and the actual risk measurements, the Laplacian MDS can provide useful indicators of influenza epidemics at the global scale. For example, a series of H1N1 outbreaks were reported by Egypt and Kenya in June 2009. Though no surveillance data were reported to the WHO, our method predicted that the epidemic risk was not negligible in nearby countries such as Morocco, Ethiopia, Algeria, Tunisia, and Mauritius, where outbreaks of the H1N1 virus were confirmed later (ECDC, 2009).

Fig. 5 compares the estimates and the actual risk measurements for all African countries. The color of the pixels in each row represents the epidemic risk for each country. For most of the African countries reporting to the WHO, their influenza surveillance data were incomplete. The black gaps in Fig. 5a illustrate the missing data, where the missing rate ranged from less than 1% to over 99% (Chad). Comparing Figs. 5a and 5b, one may see that given the data missing challenge, the estimated risk still captures the high and low points of the actual influenza epidemic risk. Since the proposed method is designed to predict the epidemic risks of unmonitored countries using the data collected from nearby countries, the risk estimates of countries in the same transmission zone showed similar patterns during the H1N1 pandemic period.

For each country or area in Africa and Asia, the correlation coefficient between the temporal sequence of estimates and the actual risk measurements is listed in Table 1. Though the Laplacian MDS was not designed to preserve temporal patterns, the correlation for the countries in Africa and Asia was relatively high. The average correlation for different African transmission zones ranged from 0.79 (Middle Africa, $\sigma = 0.05$) to 0.99 (Southern Africa, $\sigma = 0.05$). Further observation shows that the average correlation for European countries ($\rho = 0.92$, $\sigma = 0.05$) or American countries ($\rho = 0.95$, $\sigma = 0.05$) is higher than that of African ($\rho = 0.81$, $\sigma = 0.05$) or Asian countries ($\rho = 0.85$, $\sigma = 0.05$). This might be caused by the higher missing rate in African and Asian surveillance data, which resulted in less smooth trends and lower temporal correlations.

To evaluate the predicted risks for the countries with no surveillance data, we designed a leave-one-out experiment where the morbidity data of the countries examined were assumed to have been missing in risk estimation. The leave-one-out experiment examined one country at a time, and the correlation between the actual risk and the predicted risk of the countries examined was computed (Table 1). Compared with the previous experiment where all morbidity data were used for risk estimation, the average correlation coefficient of the leave-one-out experiment dropped about 13% on average for African and Asian countries. However, for all countries, a positive correlation between the predicted risk and actual risk was observed, and the correlation for 53% countries was higher than 0.70 ($\sigma = 0.05$), which

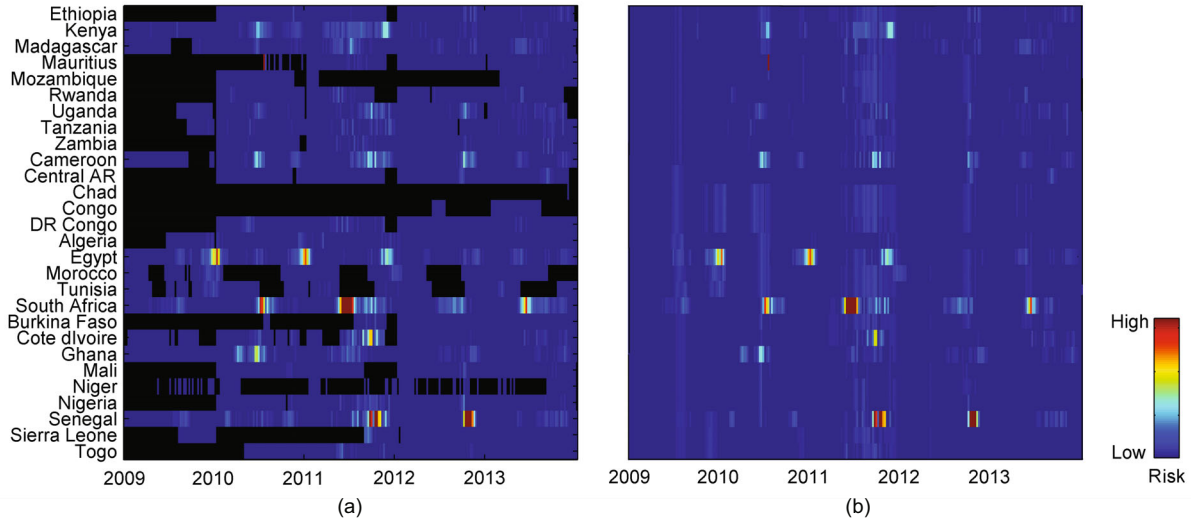

Fig. 5 The actual risk measurement (a) and the estimated risk (b) for all African countries of five influenza transmission zones (References to color refer to the online version of this figure)

Table 1 Correlation coefficients between the temporal sequence of estimates and the actual risk measurement for Asian and African countries ($\sigma = 0.05$)

African zone	Country or area	Missing rate (%)	Temporal correlation	Leave-1-out correlation
Eastern Africa	Ethiopia	23.46	0.71	0.69
	Kenya	0.00	0.90	0.74
	Madagascar	5.00	0.79	0.72
	Mauritius	40.00	0.91	0.81
	Mozambique	63.08	0.64	0.61
	Rwanda	29.23	0.73	0.62
	Uganda	15.77	0.86	0.71
	Tanzania	14.62	0.72	0.66
	Zambia	21.92	0.75	0.64
Middle Africa	Cameroon	5.00	0.91	0.79
	Central AR	25.77	0.75	0.70
	Chad	99.99	–	–
	Congo	86.15	0.63	0.52
	DR Congo	23.08	0.72	0.56
Northern Africa	Algeria	9.62	0.80	0.80
	Egypt	0.00	0.96	0.88
	Morocco	37.31	0.72	0.64
	Tunisia	33.08	0.69	0.62
Southern Africa	South Africa	0.00	0.99	0.73
Western Africa	Burkina Faso	51.54	0.65	0.65
	Cote dIvoire	14.62	0.96	0.82
	Ghana	0.00	0.93	0.79
	Mali	28.08	0.72	0.61
	Niger	67.69	0.66	0.56
	Nigeria	21.15	0.84	0.71
	Senegal	0.00	0.97	0.78
	Sierra Leone	45.00	0.73	0.67
	Togo	26.54	0.89	0.83

Asian zone	Country or area	Missing rate (%)	Temporal correlation	Leave-1-out correlation
Eastern Asia	China	0.00	0.99	0.73
	Japan	0.00	0.99	0.78
	Mongolia	0.38	0.72	0.62
	North Korea	0.38	0.94	0.89
Central Asia	Kazakhstan	23.85	0.80	0.73
	Kyrgyzstan	16.54	0.75	0.54
	Uzbekistan	82.69	0.59	0.58
Southern Asia	Afghanistan	54.23	0.72	0.52
	Bangladesh	20.38	0.95	0.82
	Bhutan	41.15	0.74	0.61
	India	5.38	0.95	0.85
	Iran	0.77	0.92	0.75
	Nepal	0.55	0.97	0.67
	Pakistan	0.00	0.81	0.74
	Sri Lanka	2.31	0.90	0.82
Southeast Asia	Cambodia	0.77	0.98	0.90
	Indonesia	20.38	0.73	0.71
	Lao PDR	37.69	0.71	0.69
	Malaysia	0.00	0.86	0.85
	Philippines	2.69	0.92	0.87
	Singapore	3.08	0.93	0.73
	Thailand	1.54	0.91	0.64
	Viet Nam	1.54	0.95	0.72
Western Asia	Armenia	36.54	0.72	0.68
	Azerbaijan	41.54	0.74	0.62
	Bahrain	51.92	0.66	0.42
	Georgia	10.77	0.83	0.72
	Iraq	0.00	0.78	0.66
	Israel	39.23	0.81	0.75
	Jordan	38.46	0.65	0.64
	Oman	5.77	0.76	0.69
	Qatar	41.92	0.82	0.77
	Turkey	13.85	0.81	0.74

indicated that the predicted risk might be a useful indicator of actual influenza risk in countries where surveillance data are not available.

To compare with the temporal-transmission models, we fitted a hidden Markov model and a Bayesian model to estimate the actual risk measurement. The Bayesian method achieved an average correlation of 0.77 ($\sigma = 0.05$) and 0.83 ($\sigma = 0.05$) for African and Asian countries, respectively, which were comparable with the Laplacian MDS. The average correlation achieved by the hidden Markov model was notably lower for African countries ($\rho = 0.47$, $\sigma = 0.05$), which was caused possibly by inadequate training data.

In the previous experiments we used the morbidity number to estimate the epidemic risks. In fact, one can also use the prevalence data, which is the fraction of infected population in a country, to estimate the influenza epidemic risk. The yearly average correlation between the estimated risk and actual epidemic risk ranged from 0.83 to 0.92 ($\sigma = 0.05$) from the year 2009 to 2013, which was high but lower than the correlation achieved using morbidity numbers. The main reason might be that the high correlation was contributed mainly by the high peaks in morbidity trends. Most peaks in the morbidity data were countries of large populations; however, the peaks in the prevalence data were relatively low, resulting in lower correlation results.

5 Discussion and conclusions

One of the most important challenges of global influenza surveillance is the lack of data (Briand *et al.*, 2014), especially in the developing countries where most epidemics emerged. To fill the surveillance gaps, we estimated the influenza epidemic risks in countries with incomplete or even no surveillance data based on the morbidity data collected in nearby countries. Experiments showed that the estimated risks were strongly correlated with the actual influenza epidemic risk in general, which could provide early warnings of potential influenza epidemics in countries lacking the infrastructure required for traditional surveillance. Moreover, for the countries with incomplete surveillance data, the proposed method could continuously provide weekly epidemic risk estimates. Though our method focuses on influenza surveillance at the global scale, the Laplacian MDS could also be extended for monitoring other infectious diseases on different scales.

The proposed method extends the coverage of disease surveillance to resource-limited countries, which is potentially very useful for disease control and prevention; however, it still has limitations. First, due to the lack of ground-truth knowledge, it is difficult to fully evaluate the risk estimates for countries with no surveillance data. Moreover, the lack of international standards for reporting risk factors and morbidity data reduces the accuracy of the estimates (Briand *et al.*, 2014). Another limitation of the Laplacian MDS is that it considers only the geographical correlation of epidemic risks. To further improve the results, we plan to combine the geographical and temporal correlations for disease surveillance in resource-limited countries.

Acknowledgements

The authors thank the anonymous reviewers and Fan YANG for their thorough comments which improve this work.

References

Best, N., Richardson, S., Thomson, A., 2005. A comparison of Bayesian spatial models for disease mapping. *Stat. Methods Med. Res.*, **14**(1):35-59. http://dx.doi.org/10.1191/0962280205sm388oa

Briand, S., Mounts, A., Chamberland, M., 2014. Challenges of Global Surveillance during an Influenza Pandemic. World Health Organization, Geneva. Available from http://www.who.int/influenza/surveillance_monitoring/Challenges_global_surveillance.pdf [Accessed on June 10, 2014].

Cooper, B.S., Pitman, R.J., Edmunds, W.J., *et al.*, 2006. Delaying the international spread of pandemic influenza. *PLoS Med.*, **3**(6):e212. http://dx.doi.org/10.1371/journal.pmed.0030212

ECDC, 2009. Pandemic (H1N1) 2009. European Centers for Disease Control, Stockholm. Available from http://ec.europa.eu/health/communicable_diseases/diseases/influenza/h1n1/index_en.htm [Accessed on June 10, 2014].

Eubank, S., Guclu, H., Kumar, V., *et al.*, 2004. Modelling disease outbreaks in realistic urban social networks. *Nature*, **429**:180-184. http://dx.doi.org/10.1038/nature02541

Ferguson, N., Donnelly, C., Anderson, R., 2001. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science*, **292**(5519):1155-1160. http://dx.doi.org/10.1126/science.1061020

Ferguson, N., Cummings, D., Cauchemez, S., *et al.*, 2005. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, **437**:209-214. http://dx.doi.org/10.1038/nature04017

Hay, S.I., Battle, K.E., Pigott, D.M., *et al.*, 2013. Global mapping of infectious disease. *Phil. Trans. R. Soc. B*, **368**(1614):20120250. http://dx.doi.org/10.1098/rstb.2012.0250

He, D., Dushoff, J., Eftimie, R., *et al.*, 2013. Patterns of spread of influenza A in Canada. *Proc. R. Soc. B*, **280**(1770):20131174. http://dx.doi.org/10.1098/rspb.2013.1174

He, D., Chiu, A., Lin, Q., *et al.*, 2015a. Differences in the seasonality of Middle East respiratory syndrome coronavirus and influenza in the Middle East. *Int. J. Infect. Dis.*, **40**:15-16. http://dx.doi.org/10.1016/j.ijid.2015.09.012

He, D., Lui, R., Wang, L., *et al.*, 2015b. Global spatio-temporal patterns of influenza in the post-pandemic era. *Sci. Reports*, **5**:11013. http://dx.doi.org/10.1038/srep11013

Hollingsworth, T., Ferguson, N., Anderson, R., 2007. Frequent travelers and rate of spread of epidemics. *Emerg. Infect. Dis.*, **13**(9):1288-1294.

Keeling, M., Woolhouse, M., Shaw, D., *et al.*, 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, **294**(5543):813-817. http://dx.doi.org/10.1126/science.1065973

Kenah, E., Chao, D., Matrajt, L., *et al.*, 2011. The global transmission and control of influenza. *PLoS ONE*, **6**(5):e19515. http://dx.doi.org/10.1371/journal.pone.0019515

Lavanchy, D., 1999. The importance of global surveillance of influenza. *Vaccine*, **17**:S24-S25.

Longini, I., Nizam, A., Xu, S., *et al.*, 2005. Containing pandemic influenza at the source. *Science*, **309**(5737):1083-1087. http://dx.doi.org/10.1126/science.1115717

Nelson, M.I., Viboud, C., Vincent, A.L., *et al.*, 2015. Global migration of influenza A viruses in swine. *Nat. Commun.*, **6**:6696. http://dx.doi.org/10.1038/ncomms7696

Oshitani, H., Kamigaki, T., Suzuki, A., 2008. Major issues and challenges of influenza pandemic preparedness in developing countries. *Emerg. Infect. Dis.*, **14**(6):875-880. http://dx.doi.org/10.3201/eid1406.070839

Riley, S., 2007. Large-scale spatial-transmission models of infectious disease. *Science*, **316**(5829):1298-1301. http://dx.doi.org/10.1126/science.1134695

Tamerius, J., Shaman, J., Alonso, W.J., *et al.*, 2013. Environmental predictors of seasonal influenza epidemics across temperate and tropical climates. *PLoS Path.*, **9**(3):e1003194. http://dx.doi.org/10.1371/journal.ppat.1003194

Wang, L., Li, X., 2014. Spatial epidemiology of networked metapopulation: an overview. *Chin. Sci. Bull.*, **59**(28):3511-3522. http://dx.doi.org/10.1007/s11434-014-0499-8

WHO, 2014. Introduction of the Influenza Transmission Zones. World Health Organization, Geneva. Available from http://www.who.int/csr/disease/swineflu/transmission_zones/en/ [Accessed on June 10, 2014].

WHO Regional Office for Africa, 2009. Pandemic (H1N1) 2009 in the African Region: Update 63. World Health Organization, Brazzaville. Available from http://www.afro.who.int/index.php?option=com_docman&task=doc_download&gid=3954 [Accessed on June 10, 2014].

Williams, C., 2002. On a connection between kernel PCA and metric multidimensional scaling. *Mach. Learn.*, **46**(1):11-19. http://dx.doi.org/10.1023/A:1012485807823

Zhou, X., Shen, H., 2010. Notifiable infectious disease surveillance with data collected by search engine. *J. Zhejiang Univ.-Sci. C (Comput. & Electron.)*, **11**(4):241-248. http://dx.doi.org/10.1631/jzus.C0910371

Zhou, X., Ye, J., Feng, Y., 2011. Tuberculosis surveillance by analyzing Google trends. *IEEE Trans. Biomed. Eng.*, **58**(8):2247-2254. http://dx.doi.org/10.1109/TBME.2011.2132132

Zhou, X., Li, Q., Zhu, Z., *et al.*, 2013. Monitoring epidemic alert levels by analyzing Internet search volume. *IEEE Trans. Biomed. Eng.*, **60**(2):446-452. http://dx.doi.org/10.1109/TBME.2012.2228264