Spatio-Temporal Bayesian Hierarchical Model for Influenza Surveillance

Ta-Chien Chan[1], Chwan-Chuen King[1], Muh-Yong Yen[2], Chuhsing Kate Hsiao[1]*

1. Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan
2. Department of Disease Control and Prevention, Taipei City Hospital, Taipei City Government, Taipei, Taiwan

Outline

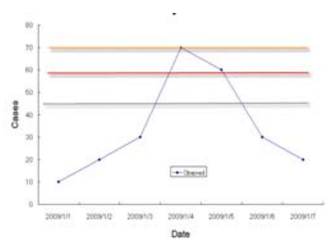
- Introduction and Rationale
  - Background of the study
  - Concept of probability expression
  - Rationale for Bayesian methods and covariates
- Aims of this study
- Methods
  - Data source
  - Model
  - Software
- Results
- Conclusion
- Discussion
- Future work

Introduction

- Characteristics of influenza
  - Short incubation period
  - Highly transmissibility
- Earlier aberration detection can save time to do prevention and intervention
- Syndromic surveillance provides timely surveillance tool
- Current mostly used methods for aberration detection
  - Historical limits
  - Cusum
  - Serfling Models
  - EWMA

Threshold vs. Probability

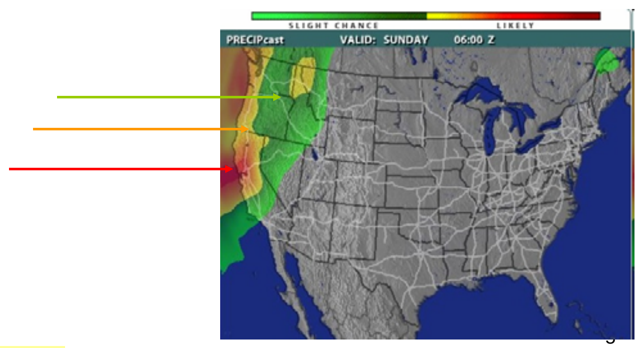
- Traditionally, the signals alerted when the prediction values exceeded the threshold <flag=1>
- Many ways to define the threshold
  - 95% Upper Confidence Interval
  - $1\sigma, 2\sigma, 3\sigma$ (SD.)
  - Average or Median values
  - Upper arm of V-mask <Cusum>



- Threshold was needed for our decision making
- How much confidence do you have in these signals ?
  - Can we express the uncertainty by probability?

Probability

- In our daily life, the probability of precipitation was a good model for disease surveillance
- Expression of probability:
  - slightly chance , Medium likely, Most likely



Why Bayesian ?

- Three advantages
  - Consider uncertainty
  - Use large posterior samples to get estimation at each time point
  - Express the chance of exceeding the threshold by probability

Bayesian methods on flu surveillance

- LeStrat and Carrat, 1999
  - Hidden Markov Model
  - Flu epidemic and non-epidemic phases
- Mugglin et al, 2002
  - Multivariate autoregressive process
  - Space-time dynamics
- Cowling et al, 2006
  - Dynamic linear model
- Sebastiani et al, 2006
  - Dynamic Bayesian Networks
- Nuno and Pagano, 2007
  - Gaussian-type regression model
- Miguel A. et al, 2008
  - Bayesian Markov switching models
- Huafeng Zhou and Andrew B. Lawson
  - EWMA smoothing and Bayesian spatial modeling
- K. M. L. Charland et al, 2009
  - Bayesian hierarchical models

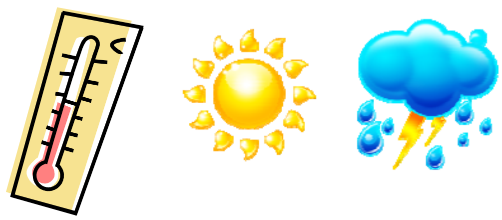
1. **Most are temporal algorithms, did not consider spatial interaction**
2. **Not many incorporated weather factors**
3. **Temporal precision is week or longer, the timeliness may be delayed**

Spatial Interaction

- Neighborhoods
  - Spatial interaction can reflect the possible patients' movement in the community
  - Assumption:
    - Within the predefined overlapped buffers, patients have multiple choices to different hospitals
    - The neighboring hospitals might have correlated effects
- Conditional autoregressive (CAR) spatial structure

Meteorological factors

- Temperature (Shaman, J. et al.,2009, Lowen, A. C. et al., 2008)
- Relative humidity (Lowen, A. C. et al., 2007)
- Vapor Pressure (Shaman, J. et al.,2009)
- Solar radiation (Charland, K. M. et al, 2009)



Aims of this study

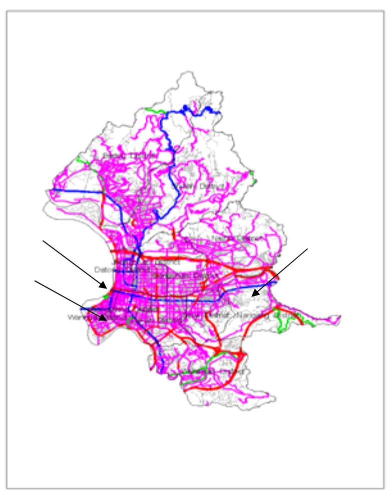
- Consider the impact of spatial and temporal dependence on influenza surveillance
- Use Bayesian hierarchical model with weather information to implement probabilistic prediction on influenza activity in Taipei City

Methods –Sources of Data

- 5 emergency rooms of community hospitals in Taipei City
- Training data
  - From the daily surveillance during Jan. 2006 – Dec. 2007
- Validation Data
  - From Jan. 2008 to Feb. 2008
- Diagnosis by ICD-9 codes
  - The influenza-like illness (ILI) was defined by the composite group of ICD-9 codes validated by the physicians
- Weather data
  - The daily meteorological data were from Taiwan Central Weather Bureau (TWCWB).

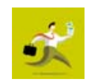
Methods –Spatial correlation

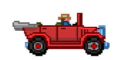
- The buffer was calculated by the real road network
- The spatial neighboring relationship was defined with a 3 km network buffer surrounding each hospital.
- E.g., hospital 3 vs. hospital 1,2,5



Transportation time from home to hospital

Types	Velocity (km/hr)	Time to reach 3 Kms (mins)
Walk	5	36
Walk	10	18
Drive/ Bus	40	4.5
Drive	50	3.6



Bayesian Hierarchical Model

- Outcome (Dependent) variable
  - ILI daily visits in each hospital
  - Poisson distribution
  - Population at risk within each buffer (census data)
- Explanatory (Independent) variables
  - Previous ILI visits
  - Mean Temperature
  - Mean Vapor pressure
  - CAR for spatial interaction
  - Weekend and holiday effect

MCMC & Converge Diagnosis

- Gibbs sampling algorithm
  - WinBUGS 1.4.3.
  - Chains=3; Samples=10,000; Burn-in: 5000; Thin=10, 15,000 samples for each parameter
- Convergence Diagnosis
  - The "coda" library of the R statistical package
  - Gelman and Rubin statistic.

Model Fitting and Prediction

- Implemented by R
- Prediction residuals→ accuracy
- Dynamic threshold
  - **Maximum value of 7 days' temporal window**
- Probability exceeding the threshold
  - By each hospital
  - By all hospitals

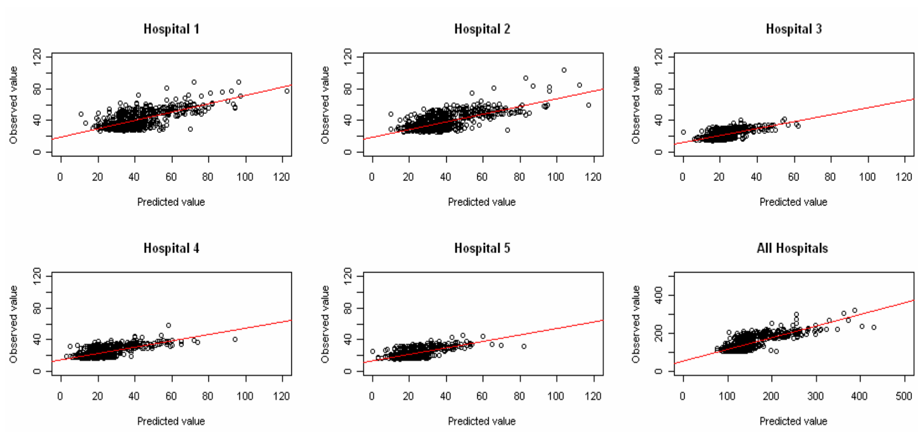
Preliminary Results

- Model fitting stage:
  - The average residual was 0.34 visits in all hospitals
  - Overall Correlation: 0.80

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	All Hospitals
Average Residual	**0.38**	**0.37**	**0.21**	**-0.82**	**0.20**	**0.34**
Correlation	**0.69***	**0.72***	**0.66***	**0.69***	**0.64***	**0.80***

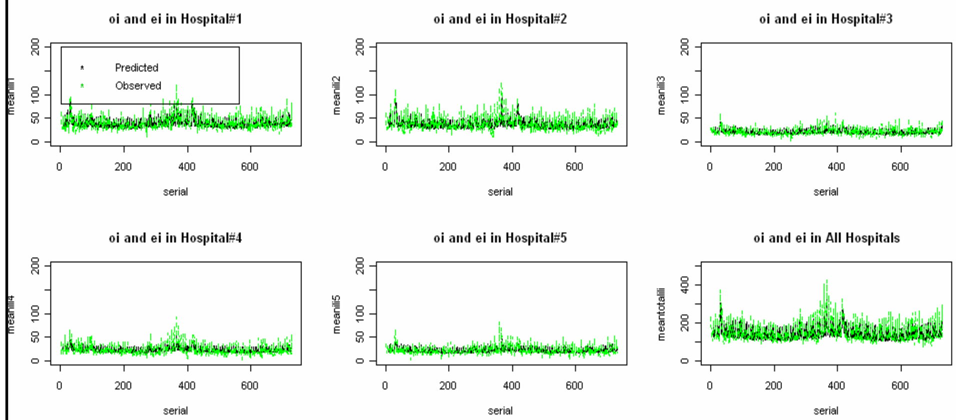
* $p<0.05$

Model fitting in each hospital



Red line: Correlation line

Temporal trends of the observed and predicted values



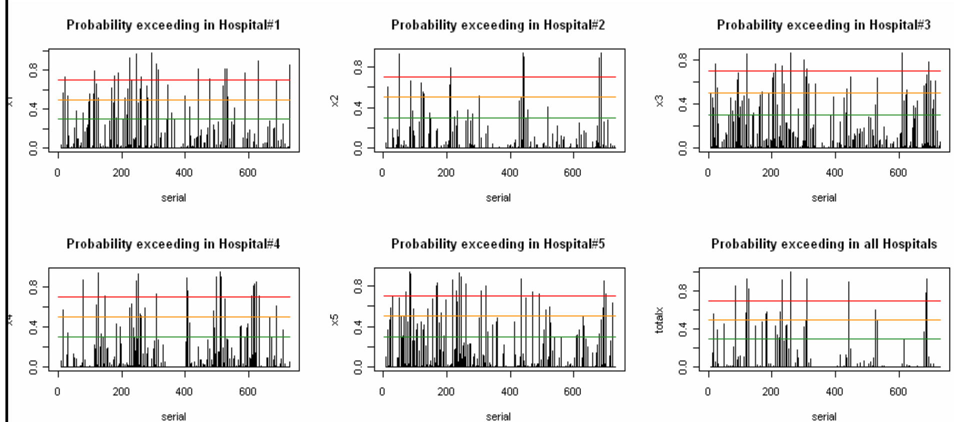
*Predicted *Observed

19

Results

Chan TC, et al.

Probability exceeding the threshold at model fitting stage



Red line: 0.7; Orange line:0.5; Green line: 0.3

20

Results

Chan TC, et al.

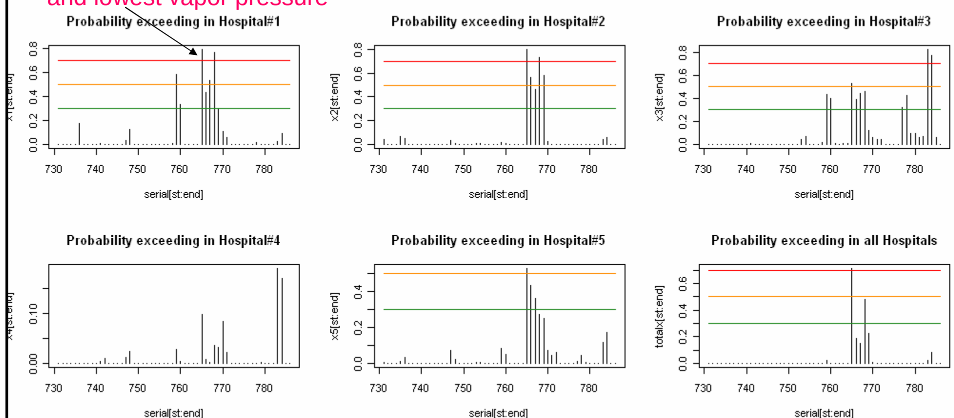
Model Validation stage (2 months)

- The average residual was -0.5 visits in all hospitals
- By hospitals:

Average Residual	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	All Hospitals
In two months	0.66	-1.47	-1.01	0.11	1.22	-0.50

Probability exceeding the threshold at validation stage

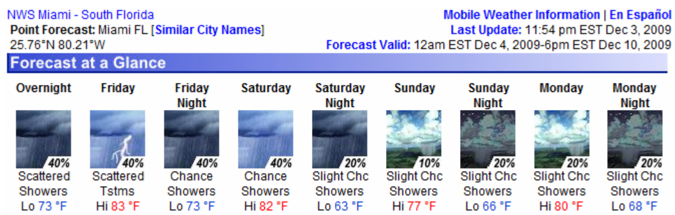
Combination effect of weekend & Start of Chinese New Year, lowest temperature and lowest vapor pressure



Red line: 0.7; Yellow line:0.5; Green line: 0.3

Conclusion

- By using the proposed method, the daily or weekly forecast of the influenza epidemic might become feasible in the community or city level



Discussion

- Holiday and weekend effects were strong
  - Most of clinics and outpatients visits were all off
- Meteorological condition might be different for different countries
- Extension for forecasting 7 days' ILI activity like weather forecast would become feasible by our model
- Sensitivity of the probability was still under testing
  - Implication of the value was important for decision making
- Limitation in validation:
  - Lack of daily influenza virus isolation data to validate the virus activity in the community level

Future direction

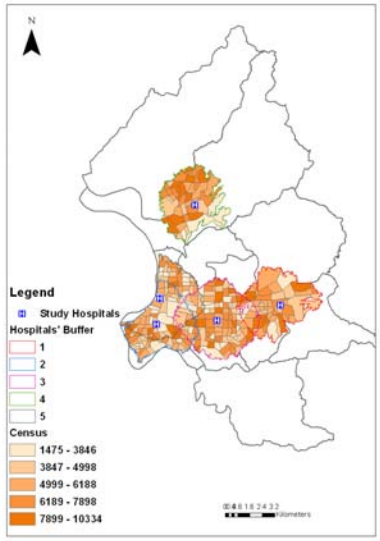
- Continues improving the accuracy of Bayesian model
- Evaluate the specificity, sensitivity and timeliness
- Make proposed method as free R's package for widely sharing

Contact information:

Ta-Chien Chan

E-mail: dachianpig@gmail.com

Population at Risk



Chan TC, et al.