Meteorological and Environmental Indicators in Seasonal Influenza Transmission

David O. Roberts
Hobart and William Smith Colleges, Geneva, NY
Mathematics 2011

Radina P. Soebiyanto
Mentor
Code 610.2 Global Change Data Center
NASA Goddard Space Flight Center, Greenbelt, MD
Goddard Earth Sciences and Technology Center
University of Maryland Baltimore County, MD

Richard K. Kiang
Code 610.2 Global Change Data Center
NASA Goddard Space Flight Center, Greenbelt, MD

August 6, 2010

1. ABSTRACT

The role of climatic factors on the spread of infectious diseases has long been acknowledged, especially on diseases such as influenza. Our group has focused on the role of meteorological factors in influenza transmission in Hong Kong. Our group has tested effects of environmental variables on influenza transmission, results for the effect of meteorological and environmental variables influenza types and subtypes will be discussed. Our models are based off of Poisson regression and Neural Network. Neural Network proves to be a more accurate model than Poisson.

2. INTRODUCTION

Influenza is an acute viral infection that spreads easily from person to person and causes yearly epidemics. Worldwide, these annual epidemics result in about three to five million cases of severe illness, and about 250,000 to 500,000 deaths (WHO, 2009). In 2009, the emergence of the influenza strain H1N1 caused immediate attention to be focused on worldwide influenza surveillance capabilities. This act was in response to the WHO determining that the world was underprepared for influenza pandemics, mostly because of the lack of surveillance (Ortiz, 2009).

In addition to the significant burden annually, influenza viruses continually change over time, usually by mutation (change in the viral RNA), that can lead to a pandemic. This constant changing often enables the virus to evade the immune system of the host so that the host is susceptible to changing influenza virus infections throughout life. This process works as follows: a host infected with influenza virus develops antibodies against that virus; as the virus changes, the "first" antibody no longer recognizes the "newer" virus and infection can occur because the

host does not recognize the new flu virus as a problem until the infection is well under way. The first antibody developed may provide some protection against infection with a new influenza virus. Unfortunately, almost all individuals have no antibodies that will recognize the novel H1N1 virus immediately. Consequently, without vaccination, the majority of the human population is susceptible to novel H1N1 flu ("Flu (influenza, conventional)," 2010).

The flu is divided into three types: A, B, and C. Among many subtypes of influenza A viruses, influenza A H1N1 and H3N2 subtypes are currently circulating among humans. Influenza types A and B are responsible for epidemics of respiratory illness that occur almost every winter in temperate regions and are often associated with increased rates of hospitalization and death. Influenza type C differs from types A and B in some important ways. Influenza Type C usually results in very mild respiratory illness or no symptoms at all; it is not responsible for pandemics and does not pose a serious threat to public-health (WHO, 2009). Influenza Type A and Type B are the main focus of this study, for they are responsible for nationwide pandemics.

This study focuses on the role of meteorological factors in influenza transmission in Hong Kong. Research has shown that low temperatures have been linked to high levels of seasonal influenza (Lofgren, 2007). Other factors that have been implicated in influenza transmission are listed in Table 1. Crowding has been known as a potential risk factor in the spread of disease. With this knowledge, it is no wonder why people contract influenza usually during the winter, when everyone huddles indoors away from harsh weather conditions. But then one may wonder why people at crowded resorts in tropical areas do not contract the viral disease. Hong Kong has a dense population, which has more potential for person-to-person spread of virus-laden aerosol particles. Hong Kong is also a subtropical region that experiences different seasons that significantly vary in temperature. Hong Kong is an advanced economy situated in the middle of the influenza basin in China, has established a sentinel surveillance network for Influenza like Illness (ILI) in the late 1990s and began reporting in 1998 (Cowling, 2006). This is a prime region for gather large amounts of data on ILI and actual laboratory confirmed cases of influenza.

Our group has tested the effects of environmental and meteorological variables on influenza transmission in Hong Kong and two cities in the US (Soebiyanto et al. (2010)). However, the differential effect of environmental and meteorological variables on influenza types and subtypes has yet to be investigated. In some regions, the differences in temporal sensitivity of influenza A and B types can be distinct. Thus understanding the differential sensitivity of influenza types and subtypes to the climactic factor will enable a more accurate prediction of influenza cases. This in turn may lead to reducing influenza burden as well as helping the influenza preparedness efforts. Toward this end, this study is aimed at modeling influenza type A and B, as well as subtypes A (H1N1) and A (H3N2) using meteorological and environmental factors. The tasks involved include collecting and processing the data needed, as well as developing linear and nonlinear models for influenza.

Table 1. Factors involved in influenza transmission

Influenza Process	Factors	Impact on influenza process from factor increase
Virus Survivorship	Temperature	↓
	Humidity	↓
	Vapor Pressure	↓
	Solar Irradiance	↓
Transmission Efficiency	Temperature	↓
	Humidity	Varies
	Vapor Pressure	↓
	Rainfall	↑
	ENSO	↑
	Air travels and holidays	↑
Host susceptibility	Sunlight exposure	↓
	Nutrition	Varies

3. MATERIALS AND METHODOLOGY

Environmental and meteorological data was obtained from ground based stations and satellite-derived measurements. We obtained data from the Hong Kong Observatory (2003), which gave us daily measurements of total rainfall, wind direction, wind speed, air pressure, air temperature, dew point, and relative humidity. In addition to ground station measurements, satellite-derived data were also retrieved, including rainfall as measured from the Tropical Rainfall Measuring Mission (TRMM) and Land Surface Temperature (LST) from MODerate resolution Imaging Spectroradiometer (MODIS). TRMM is a collaborative mission between NASA and Japan Aerospace Exploration Agency working to monitor and study the tropical rainfall so as to improve understanding of water cycle in the climate system. Data was obtained from the Tropical Rainfall Measuring Mission (TRMM) Online Visualization and Analysis System (TOVAS), all of which contained information for daily, weekly, and monthly accounts of rainfall factors. The groups collected data from LST prior to this study. Weather anomalies were further calculated in order to get a better idea of the trend of the weather patterns. The environmental variables were also smoothed using averages over the span of five points.

Influenza data were obtained from The Centre for Health Protection (2008), which summarizes and monitors the weekly situation of Human Swine Flu and seasonal influenza in Hong Kong. Because seasonal infection rates of influenza differ in each region, there is no standard number of specimens to be collected and studies from each site. A hospital surveillance system is the most effective way to acquire data and specimens that have the influenza-like illness. A specimen with an influenza-like illness one with a sudden onset of a fever exceeding 38 °C, cough or sore throat, and no other symptoms. “As part of the International Health Regulations 2005 core surveillance and response capacity requirements, each Member State must develop and maintain capabilities to detect, assess, and report disease events nationally and internationally to WHO within 48 hours of confirmation” (Ortiz, 2009). In Hong Kong, the

number of specimens tested was available, so we were able to use the proportion of positive isolates, which is the total specimens tested in that span of time, as part of our analysis.

The sentinel surveillance data from Hong Kong are provided by a network of General Practitioners (GP) and General Outpatient Clinics (GOPC). At the end of each week, the sentinel practitioners report the number of consultations with patients with ILI as well as total number of consultations (Cowling, 2006).

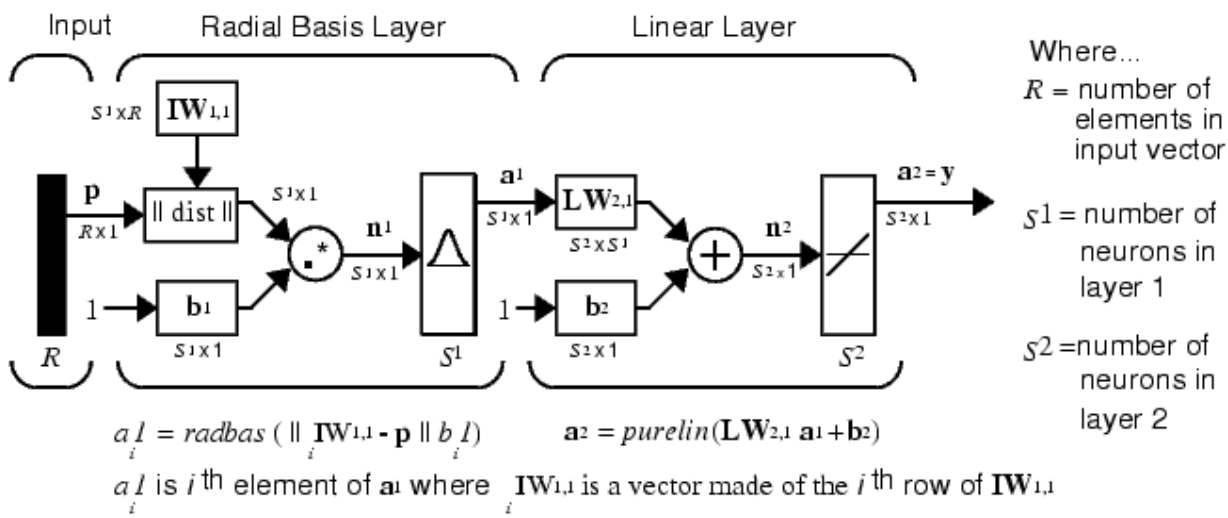
The goal of this project is to model influenza transmission rates using weather information and predictions. When analyzing the data, we put it through a series of tests. The tests that we administered were: correlation, linear regression, and Neural Network. First we found out if any of the environmental variables had any significant linear and nonlinear correlation with the flu time series. Finding correlation coefficients helps us measure of the strength of linear relationship between variables. When there is no correlation between the two quantities, then there is no tendency for the values of one quantity to increase or decrease with the values of the second quantity.

Once we found that their existed a linear relationship between the variables and influenza, the next stage was to test Poisson linear regression, which assumes

$$E(y_t|x_t) = \exp\left(\beta_0 + \sum_{i=1}^{N} \beta_i x_{ti}\right)$$

Where x_t is the column vector of independent variables on day t with regression coefficients β and y_t is the dependent variable on day t (Tobias, & Saez, 2004). Linear regression has many practical uses. We used linear regression to fit a predictive model to an observed set of y and x values, in this case, to find a fit between environmental variables and influenza ("Linear Regression," 2010). Building off of the correlation tests, we picked the environmental variables with the highest correlation to the respective strain of influenza (A or B), and calculated the least squares fit. Calculating least squares fit is useful because the least squares approach can be used to fit models that are not linear. However, because the data was surveillance information, it tended to fluctuate occasionally, leaving us with some inaccurate information, so by smoothing the environmental variables, we were able to see the trend of the transmission rates during each season.

Once we thoroughly tested our data using linear regression, we moved on and influenza incidences were modeled and predicted using Neural Network (NN). In a short description of NN, it is a process that mimics the networking function of the brain. This artificial brain process is made up of interconnecting nodes, which can be used for solving problems without necessarily having to create model of a real system by carrying a threshold function such as a sigmoid function (Soebiyanto & Kiang) (Picture 1).



Picture 1. Neural Network conceptual schematic (from MATLAB)

The nodes are arranged in layers, which are the input, hidden, and output layers. The input layer takes the data, applies parameters and weights to the data, and calculates the threshold function. The results from the first layer are then sent through the same process again and again depending on the number of nodes in the system.

In particular, we used the Radial Basis Function Neural Network (RBFNN). RBFNN is used for various applications, such as function approximation (Soebiyanto & Kiang). When using RBFNN, we limited the number of nodes to 3 in order to minimize the use of parameters and to simplify the model. Our inputs for RBFNN included the lagged environmental variables up to lag 3, as well as the previous year's data. The resulting NN model was analyzed and evaluated based on the Root Mean Square Error (RMSE) and the best correlation coefficient (R^2). All calculations, models, and simulations were tested using MATLAB. The data is divided into fit and prediction.

4. RESULTS

Our first attempt to finding correlations and determinants for influenza involved using Poisson Linear Regression. We tested weekly gathered data first. After running tests to verify correlation between different environmental variables and the types of influenza, we sorted the most highly correlated variables and ran simulations to find out if any of them were possible determinants. The combination of variables with the best root mean square error fit for influenza type A contained: solar radiation, dew point mean, mean relative humidity, minimum relative humidity, all with lag 1, and information on transmission rates and surveillance from 3 months prior. Of all the tests and simulations, relative humidity variable seems to be the determinant for Flu A (*Appendix*, Table 3).

When testing influenza Type B, the best combination of variables with the best RMSE fit is: land surface temperature, mean air pressure, maximum temperature, minimum temperature,

all with lag 2, and information regarding surveillance information from the 3 months prior (*Appendix*, Table 3). Minimum temperature seems to be the determinant for Flu B.

After testing for Flu A and B, we tested GOPC-Influenza-Like-Illness (ILI) rate, per 1000 consultation, from General Outpatient Clinics-and GP-ILI rate from General Practitioner. These are the records of all influenza like illnesses, so the reports are not confirmed cases of influenza; they are the records of patients with the symptoms before the lab results returned positive or negative. Results for GOPC reported that land surface temperature (lst) seemed to be the determinant. However, these results are from data that yielded lower levels of correlation between GOPC and the environmental variables. It seems that lst has an influence on illness, but these are not confirmed cases of influenza. When testing GP, lst again seemed to be the determinant (*Appendix*, Table 3).

The next step in our testing was to smooth the variables in order to see a better trend in our information. We performed the same correlation tests and Poisson Regression with better results.

Taking from the most correlated environmental variables, with inputs limited to 5, except for GP since there are only 3 most correlated variables, minimum relative humidity and land surface temperature seemed to be determinants for influenza Type A. The mean relative humidity and land surface temperature seem to be a determinant for influenza B, and land surface temperature and rainfall appear to be determinants for GOPC and GP (*Appendix*, Table 4).

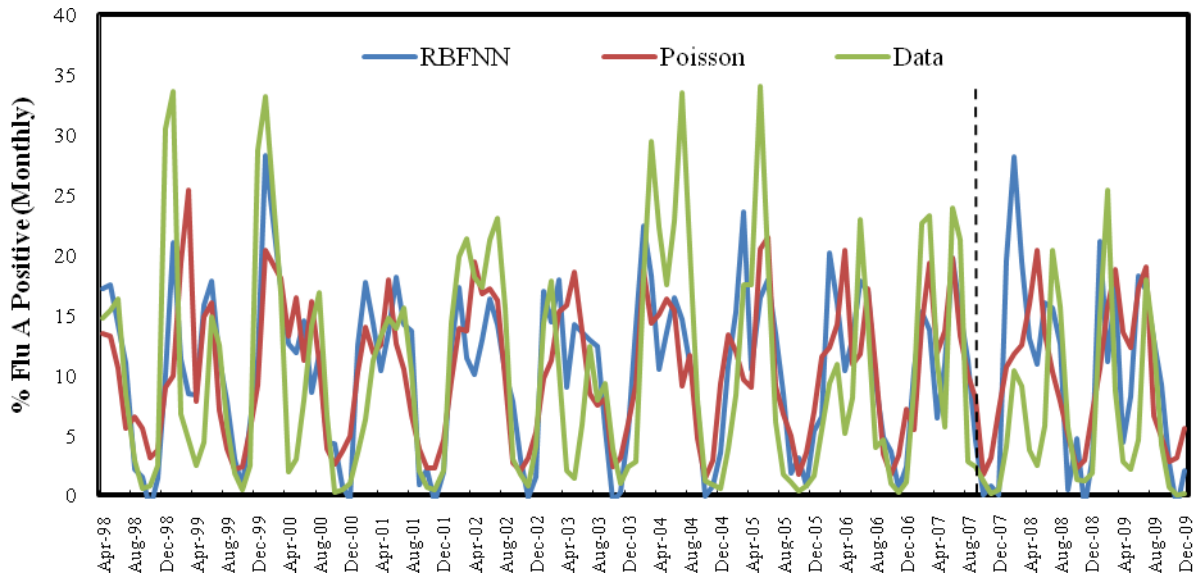
Still trying to get a clear trend of the data, we decided to test the proportion of influenza types and total specimens collected. The data we have is only from surveillance, and it doesn't reflect the true population data, so we divided the influenza type by the number of total specimens, dubbing the influenza types A PROP, B PROP, and POS PROP, where POS PROP is the combined confirmed cases of both influenza types A and B. Then we ran Poisson Linear Regression simulations. However, it did not seem as though these results were any better than the smoothed variables simulations. Determinants for A PROP seemed to be mean air pressure, mean and maximum temperature (Figure 1). Determinants for B PROP seemed to be minimum temperature and radiation (Figure 2). And for POS PROP, the determinant seemed to be minimum temperature (*Appendix*, Table 5).

After finishing the regression simulations without convincing results, we turned to Neural Network simulations for better answers. After programming the radial basis function neural network to run through all possible combinations of 3 variables, with varying spread and a maximum of 3 to 4 nodes, the outcome had a better RMSE fit all around. We kept the maximum nodes to no more than 4 in order to limit the number of parameters in the model and to keep the model as simple as possible. For these neural network simulations, we are still using weekly data for A, B, GOPC, and GP; monthly data for A PROP, B PROP and POS PROP; as well as H1N1 and H3N2.

A PROP seemed to have wind direction as its determinant, however, there is no literature confirming that wind direction has caused long distance transmission of influenza, usually the

virus dies while on course. So we then did the same process but excluding wind direction, the results were not as good, the RMSE rose by about .3 and the R^2 dropped by about .04, but the determinant seemed to be the mean dew point (Figure 1).

Figure 1



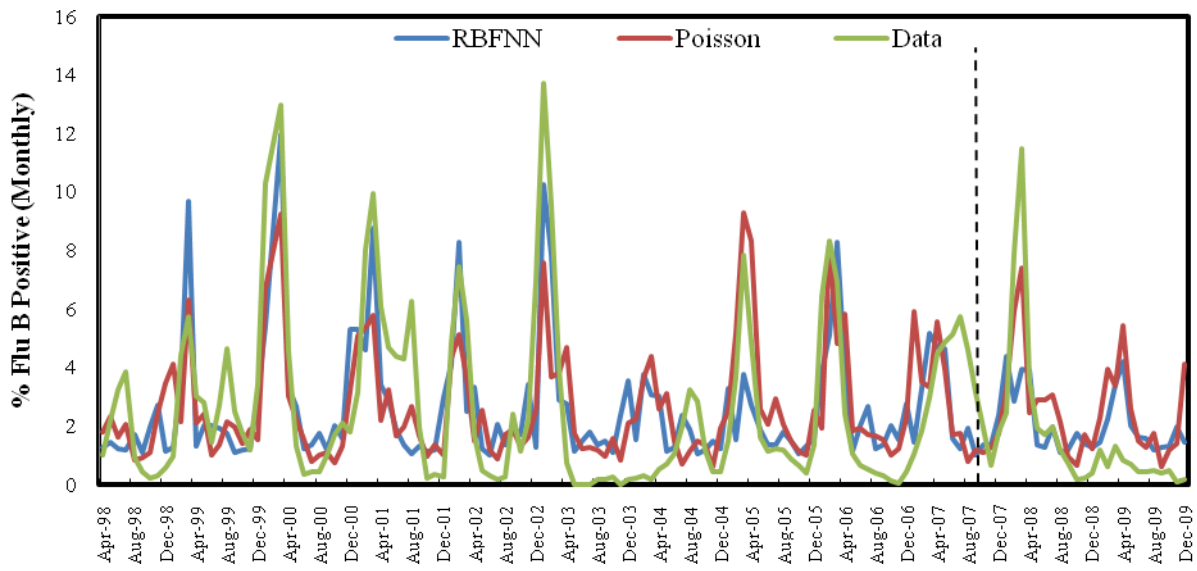
The other variables contributing to the RMSE and R^2 values are total rainfall (lag 3) and minimum temperature (lag 2) (*Appendix*, Table 6). As mentioned earlier, temperature has been directly related to aerosol-borne transmission. Rainfall has not been proved to directly affect transmission rates of influenza; however, some papers have noted that a number of tropical countries observed influenza transmission peaks that coincide with the rainfall season. Both cold temperatures and rain can also cause a person to stay indoors, which causes people to crowd together, which makes the transmission of influenza much more probable.

From all the tests on B PROP, there seemed to be no variable that was consistent with each test, so there seemed to be no determinant for B PROP (Figure 2) (*Appendix*, Table 6). However, even though it was not one of the three best variables in each test, minimum temperature (lag 1) was the most consistent variable that seemed to affect the model. This variable could be a direct cause of influenza B transmission. Studies have shown that low temperature and humidity have been most influential in aerosol-borne transmission.

Wind direction and total rainfall seemed to be the determinants for POS PROP. However, because there is not much data proving that wind direction has an effect on transmission rates, we tested the combinations again without wind direction. The results were not as promising, RMSE increase by .4 and R^2 decreased by .6. Yet total rainfall was still relatively consistent, so total rainfall seems to be a determinant of POS PROP (*Appendix*, Table 8). This makes sense

because POS PROP is the combination of confirmed cases of influenza type A and B, and A was influenced by rainfall (*Appendix*, Figure 3).

Figure 2

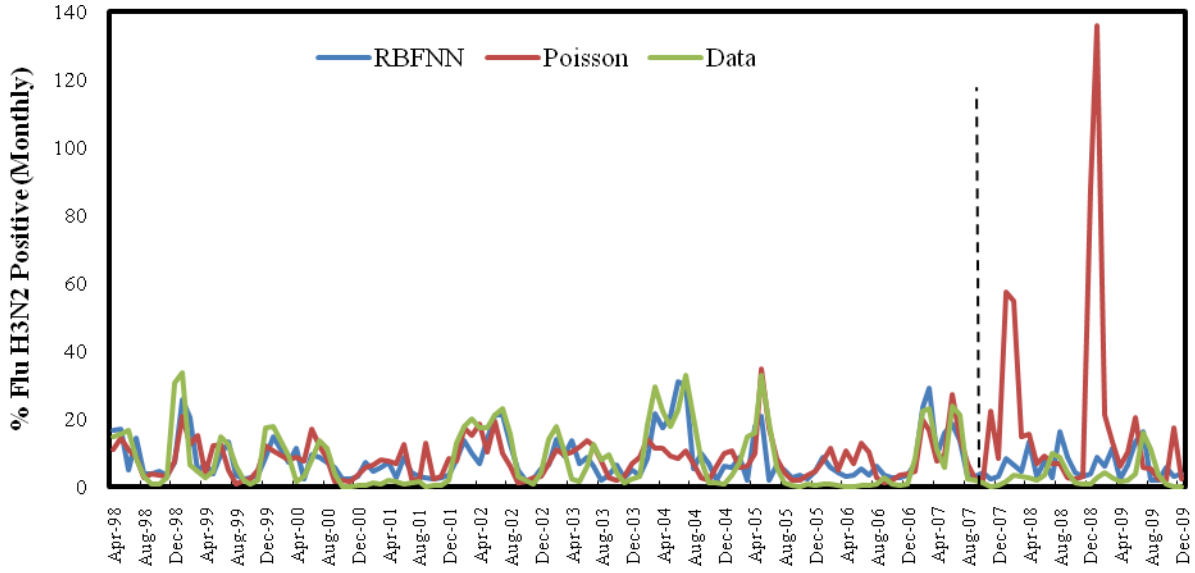


When testing cases of H1N1 proportion, mean relative humidity anomaly, sun (lag 1) and mean air pressure (lag 3) seem to be determinants. For H3N2, wind direction, sun (lag 1), and mean air pressure (lag 2) seemed to have the best RMSE fit, but when adding H3N2 (lag 1) to the test variables, the new determinants seemed to be sun (lag 1), air pressure (lag 3), and H3N2 (lag 1), which also produced a better fit (Figure 5).

When testing weekly data for influenza type A, the best combination with the lowest RMSE and highest R^2 values is: max temperature, Information for Type A from the previous year (A_t52), and Information for A from the previous 3 weeks (A_t3). However, results for influenza Type A tests had sporadic results, thus it seems as though there is not just one environmental determinant. When testing weekly data for influenza type B, the best combination of variables is: mean air pressure (lag 2), min temperature (lag 3), and Information for Type B from the previous 3 weeks (B_t3). Each test seemed to favor some form of temperature as one of the determinants, whether it was max, min, with lag 3 or lag 2.

Since GOPC had the most hits, it was simulated using RBFNN rather than GP. The best combination of variables concerning GOPC is: mean relative humidity, mean relative humidity with lag 3, and min temperature with lag 2. GOPC did not seem to have any determinants.

Figure 3



5. DISCUSSION

Influenza is a detrimental virus that causes yearly pandemics, and a pandemic is determined by spread of disease, not its ability to cause death. Influenza's ability to spread easily from person to person can cause serious public health and economic problems. In developed countries, epidemics can result in high levels of productivity losses. In communities, clinics and hospitals can be overwhelmed when large numbers of sick people appear for treatment during peak illness periods. While most people recover from a bout of influenza, there are large numbers of people who need hospital treatment and many who die from the disease every year (WHO, 2009).

The information found out from our Linear Regression and Neural Network simulations have given us new insights into direct environmental causes of influenza. It seems as though wind direction may have some sort of indirect link to the transmission of the disease, even though there may not be enough research done to support that hypothesis. Relative humidity, temperature and land surface temperature seem to be the most consistent with all types of influenza and probabilities. Tests have been done to tests whether or not humidity has a direct effect on transmission rates, the results came back positive. Now that we have more information on which variables may influence transmission of influenza, we can do more direct tests to confirm our assumptions.

For all of the simulations, the neural network model does the best job mimicking the data and creating forecasts and predictions. For influenza type A, our neural network model of the monthly data was the most accurate. This model used mean dew point (lag 0), minimum temperature (lag 2) and total rainfall (lag 3) as its determinants. For influenza type B, the models using monthly data instead of weekly data proved to be more accurate. For the most part, the neural network model for Flu B, which used wind speed (lag 0), maximum temperature (lag 1),

and Flu B (lag 2), was accurate when studying and creating a pattern for the data, but was not very promising when it forecast predictions. Neural network models seem more accurate than the Poisson linear regression models for POS PROP, H1N1, H3N2, and GOPC; however these models are still a work in progress.

The more data collected for this research, the more accurate the models will become. If the sentinel surveillance networks continue to be installed globally then we will have more access to accurate and numerous results and information on cases of influenza. If we continue to research how environmental factors individually affect influenza transmission, such as previous work done with humidity and temperature, we will be more efficient in forecasting and predicting future outbreaks.

Appendix

Table 2

Poisson Regression Weekly Data					
Output	**Input**	**RMSE Fit**	**RMSE Pred.**	**R^2 Fit**	**R^2 Pred.**
A	radiation (lag 1) mean dew point (lag 1) mean relative humidity (lag 1) min. relative humidity (lag 2), Flu A (lag 3)	67.80277	92.33076	0.4385	0.01538
B	land surface temp. (lag 2) mean pressure (lag 2) max. temp. (lag 2) min. temp. (lag 2) Flu B (lag 3)	15.73687	218.7049	0.54237	0.03051
GOPC	rainfall (lag 1) land surface temp. (lag 1) sunlight (lag 1) radiation (lag 1) GOPC (lag 3)	1.41874	1.97455	0.34508	0.12973
GP	land surface temp. (lag 1, 2) max. temp. (lag2) GP (lag 52, 3)	6.95685	9.76129	0.45611	0.06407

Table 3

Poisson Regression Weekly Data Smooth					
Output	**Input**	**RMSE Fit**	**RMSE Pred.**	**R^2 Fit**	**R^2 Pred.**
A	rainfall (lag 1), land surface temp. (lag 1) evaporation (lag 1) max. temp. (lag 1) min. relative humidity (lag 1)	49.33816	86.0695	0.71111	0.00932
B	rainfall (lag 1) land surface temp. (lag 1) cloud coverage (lag 1) radiation (lag 1) Flu B (lag 3)	13.72931	171.2856	0.64729	0.04575
GOPC	rainfall (lag 1, 2) land surface temp. (lag 1) radiation (lag 1) GOPC (lag 3)	1.1647	2.0948	0.55886	0.1991

GP	rainfall (lag 1, 2) land surface temp. (lag 1, 2) GP (lag 3)	6.45753	13.42246	0.53147	0.00134

Table 4

Poisson Regression – Monthly Data					
Output	**Input**	**RMSE Fit**	**RMSE Pred.**	**R^2 Fit**	**R^2 Pred.**
A PROP	mean pressure (lag 0, 2) max. temp. (lag 0, 2) radiation (lag 0)	7.54556	7.42197	0.31317	0.17642
B PROP	min. temp. (lag 1) max. temp. (lag 2) mean dew point (lag 1, 2) radiation (lag 2)	2.08084	1.96701	0.47576	0.41211
POS PROP	mean pressure (lag 0, 2) mean temp. (lag 0) min. temp. (lag 0) evaporation (lag 0)	8.06854	9.12282	0.36819	0.1447

Figure 4

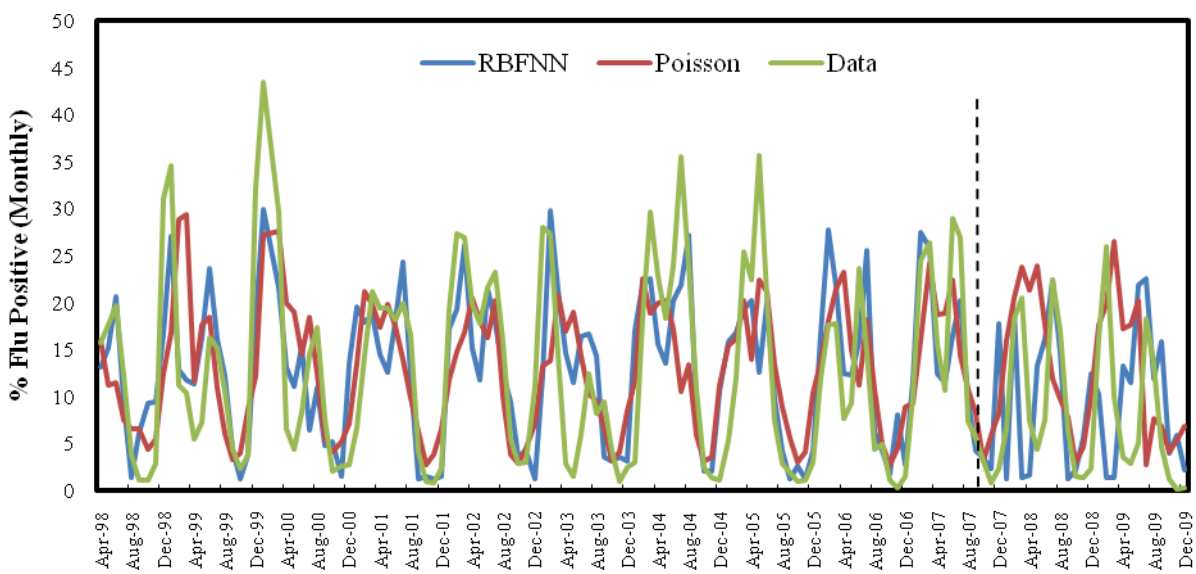


Figure 5

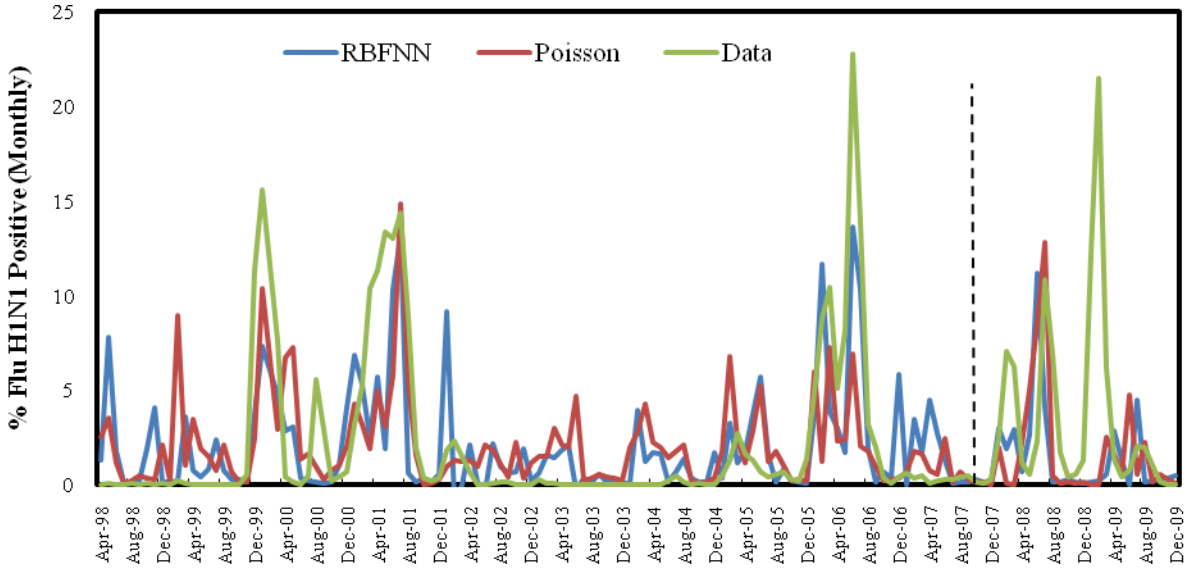


Figure 6

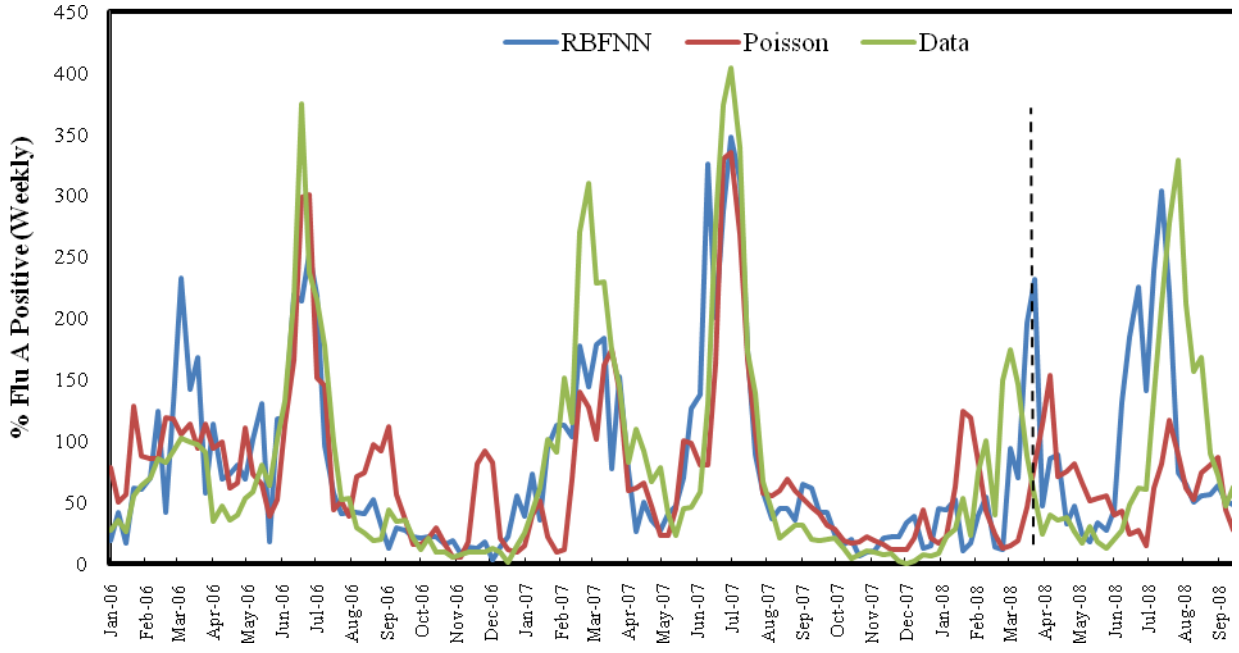


Figure 7

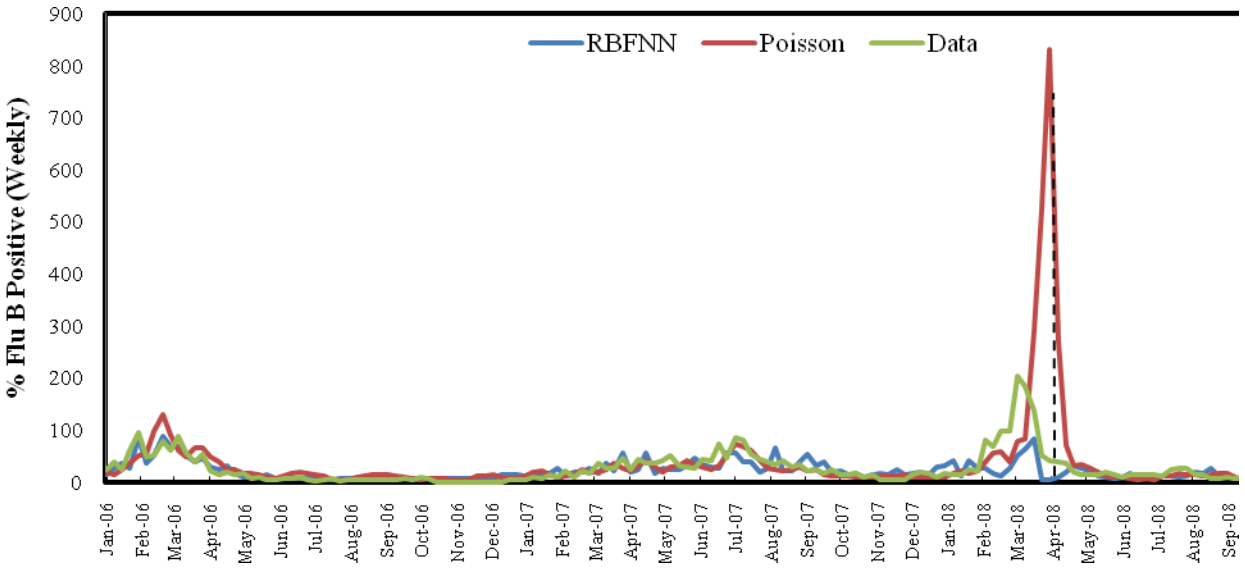
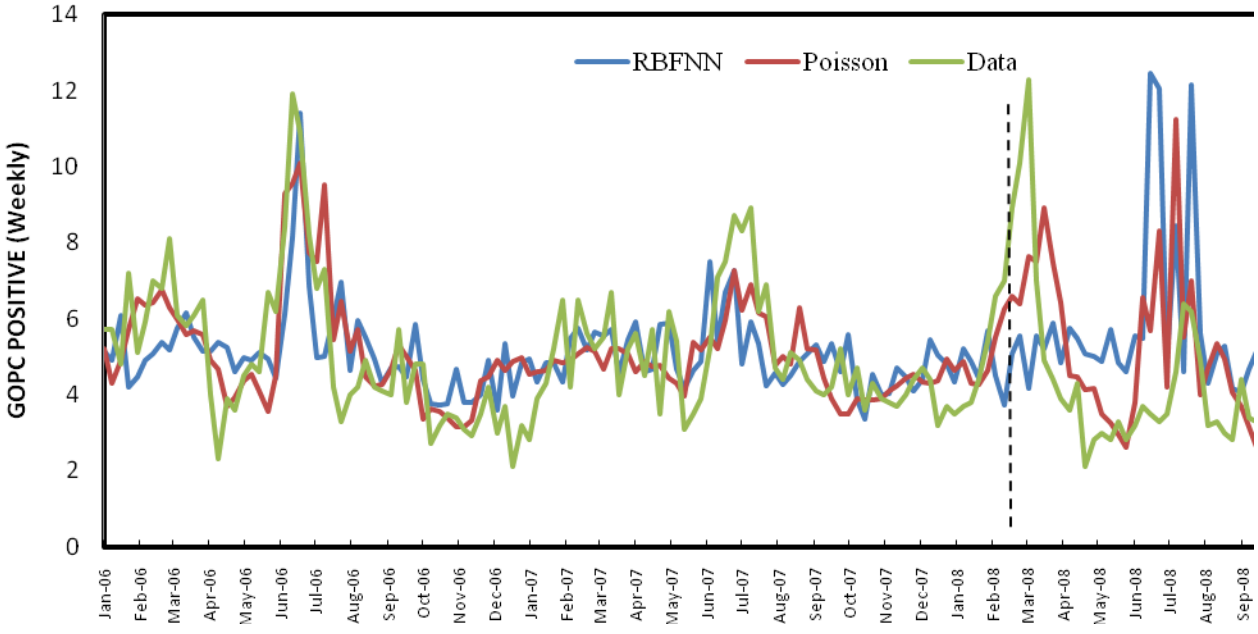


Figure 8



References

Centre for health protection: flu express. (2008). Retrieved from http://www.chp.gov.hk/en/guideline1_year/29/134/304.html

Cowling, B. J. (2006). Methods for Monitoring Influenza Surveillance Data. *International Journal of Epidemiology*, *35*, 1314-1321.

Flu (influenza, conventional and h1n1). (2010). Retrieved from http://www.medicinenet.com/influenza/article.htm

Hong Kong Observatory: extract of automatic weather station. (2003). Retrieved from http://www.weather.gov.hk/cis/data/awsext_e.htm

Linear Regression. (2010). *Wikipedia*. Retrieved (2010, July 14) from http://en.wikipedia.org/wiki/Linear_regression

Lofgren, E. (2007). Influenza Seasonality: Underlying Causes and Modeling Theories. *Journal of Virology*, *81*(11), 5429-5436.

NOAA Satellite and Information Service. (2010, July). Retrieved from http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv=&georegionabbv=

Ortiz, J.R. (2009). Strategy to Enhance Influenza Surveillance Worldwide. *Emerging Infectious Diseases*, *15*(8), 1271-1278.

Soebiyanto, R. P., & Kiang, R. K. Climate and influenza: a review.

Tobias, A., & Saez, M. (2004). *Time-series regression models to study the short-term effects of environmental factors on health.*

TRMM Online Visualization and Analysis System (TOVAS). (2010, July). Retrieved from http://disc2.nascom.nasa.gov/Giovanni/tovas/

World Health Organization (WHO) (2009) Influenza (Seasonal). Available: http://www.who.int/mediacentre/factsheets/fs211/en/index.html