

MODELING INFLUENZA TRANSMISSION USING ENVIRONMENTAL PARAMETERS

R. P. Soebiyanto^{a, b}, R. K. Kiang^{a*}

^a Global Change Data Center, NASA Goddard Space Flight Center, Code 610.2, Greenbelt, Maryland USA - (radina.soebiyanto, richard.kiang}@nasa.gov

^b Goddard Earth Sciences and Technology Center, University of Maryland Baltimore County, Baltimore, Maryland USA

Commission VIII, WG2

KEY WORDS: Influenza, modelling, environment

ABSTRACT:

Influenza is an acute viral respiratory disease that has significant mortality, morbidity and economic burden worldwide. It infects approximately 5-15% of the world population, and causes 250,000 – 500,000 deaths each year. The role of environments on influenza is often drawn upon the latitude variability of influenza seasonality pattern. In regions with temperate climate, influenza epidemics exhibit clear seasonal pattern that peak during winter months, but it is not as evident in the tropics. Toward this end, we developed mathematical model and forecasting capabilities for influenza in regions characterized by warm climate – Hong Kong (China) and Maricopa County (Arizona, USA). The best model for Hong Kong uses Land Surface Temperature (LST), precipitation and relative humidity as its covariates. Whereas for Maricopa County, we found that weekly influenza cases can be best modelled using mean air temperature as its covariates. Our forecasts can further guides public health organizations in targeting influenza prevention and control measures such as vaccination.

1. INTRODUCTION

Despite vaccination and other prevention and control efforts, influenza mortality, morbidity and economic burden remains significant. Worldwide and annually, influenza infection is between 5 – 15% of the population, and mortality can reach up to 500,000 (WHO, 2009). Whereas in the US, annual influenza epidemic causes up to 200,000 hospitalization and more than 30,000 deaths (CDC, 2010), with estimated economic burden of US\$87.1 billion (Molinari et al., 2007). In regions with temperate climate, influenza epidemics exhibit clear seasonal pattern that peak during winter months, but it is not as evident in the tropics (Finkelman et al., 2007; Viboud et al., 2006). This latitudinal variation often suggests the role of environmental factors in influenza transmission, though there is yet any definitive relationship or biological explanation underlying the seasonality pattern. Furthermore, studies that have assessed the contribution of environment to influenza seasonality and transmission – biologically and empirically – mostly arrived at region-specific results that frequently only explain the wintertime peak in temperate regions. Toward this end, we developed mathematical model and forecasting capabilities for influenza in regions characterized by warm climate. Our forecasts can further guides public health organizations in targeting influenza prevention and control measures such as vaccination.

Influenza can be transmitted via direct contact, indirect contact through contaminated objects, as well as droplets and aerosols resulting from coughs or sneezes. Any of these transmission pathways can be modulated by environmental factors. In additions, environments also plays role in the virus survivorship and the hosts susceptibility. Factors frequently associated with influenza process – transmission efficiency, virus survival and

host susceptibility – are summarized in Table 1. The table also indicates whether the relationship was revealed through biological experiments or empirical study.

Several studies have revealed that cold and dry condition, as characterized by low and temperature and humidity, provides the most effective condition for influenza virus survival and transmission (Chan et al., 2009; Lowen et al., 2007; NIH/National Institute of Child Health, 2008; Polozov et al., 2008; Shaman and Kohn, 2009; Shaman et al., 2009; Urashima et al., 2003). This supports the high influenza incidences in the winter for temperate regions.

The tropics, on the other hand, are characterized by less variability in both temperature and humidity, and hence these factors seem to have no critical role on influenza transmission in the region. An *in vivo* experiment indeed has established evidence showing that high temperature (30°C) blocks aerosol-borne transmission, but not contact transmission (Lowen et al., 2008). The latter implies that direct contact transmission is the prevailing pathway for influenza spread in the tropics. A meteorological factor that is frequently associated with high influenza incidences in the tropics is rainfall. Though the biological relationship with influenza transmission and virus survival remains to be elucidated, it increasingly becomes a common thread in the tropical countries such as Brazil (Moura et al., 2009), Singapore (Chew et al., 1998) and Thailand (Chumkiew et al., 2007). It is thought that rainfall promotes crowding tendency – as people tend to stay indoor on rainy days – that in turn increases the risk for contact transmission.

Another factor that has been associated with influenza is sunlight's ultraviolet radiation, as it possibly increases the inactivation of viruses in the environment (Sagripanti and Lytle, 2007). Furthermore, sunlight is linked to host's susceptibility to

* Corresponding author.

influenza. Human susceptibility to influenza depends on the protective capability of the immune system against the virus. The functioning of immune system, in turn, requires melatonin and Vitamin D that is determined by photoperiod (Lofgren et al., 2007). According to Dowell (2001), light-dark cycles regulate the hormone melatonin, and the amount of sunlight directly regulates Vitamin D level. The role of sunlight is further assessed through a mathematical model which evidently showed that the epidemic timing of influenza-related pediatric hospitalization in 35 cities of the US is significantly associated with solar radiation (Charland et al., 2009).

Influenza process affected	Factors (Increase in ...)	Impact on influenza process as revealed by	
		Biological	Empirical
<i>Virus survivorship</i>	Temperature	↓	
	Humidity	↓	
	Vapor pressure	↓	
	Solar irradiance		↓
<i>Transmission efficiency</i>	Temperature	↓	↓
	Humidity	varies	varies
	Vapor pressure	↓	↓
	Rainfall		↑
	ENSO		↑
	Air travel		↑
<i>Host susceptibility</i>	Holidays		↑
	Sunlight exposure	↓	
	Nutrition	varies	varies
	Previous infections	↓	↓

Table 1 Factors influencing influenza. Biological and Empirical in the last two columns refer to whether the relationship was revealed through biological experiments or empirically. Down arrow indicates reduced influenza process affected as a result of an increase in the corresponding factor.

Studies that model influenza using environmental factors primarily utilize empirical methods – such as regression – since the underlying environmental-influenza relationships remains elusive (Chew et al., 1998; Liao et al., 2009; Moura et al., 2009; Tang et al., 2010; Urashima et al., 2003; Youthao et al., 2007). Among the very few biological-based influenza model that uses environmental parameters is the compartmental Susceptible-Exposed-Infected-Recovered model where absolute humidity is included in the transmission rate (Shaman et al., 2009). In this study we employ classical time series regression technique, Autoregressive Integrated Moving Average (ARIMA) to model and forecast seasonal influenza in Hong Kong and Maricopa County (Arizona). Details on the modelling approach as well as the data are given in the next section.

2. MATERIALS AND METHODS

Hong Kong (China) lies in the tropics with centre latitude of 22°N, whereas Maricopa County (Arizona, USA) is in the subtropics and centre latitude of 34°N. Both regions are characterized by warm climate where average winter

temperature rarely goes below 0°C. In the summer, hot and humid condition prevails in Hong Kong with temperature that can go beyond 30°C during the day. Maricopa County is much less humid compared to Hong Kong, and the average summer high is about 41°C.

2.1 Influenza Data

Influenza surveillance data was obtained from the respective public health department report available publicly (Hong Kong Department of Health, 2009; Maricopa Department of Public Health, 2009). Specifically, we retrieved weekly count of lab-confirmed influenza positive for each region. The weekly data spans January 2005 – September 2008 for Hong Kong, and October 2004 – March 2009 for Maricopa County. Observations from the latest year were reserved to validate our forecast, while the remaining was used in the fitting process during which the parameters were estimated. Since Maricopa County influenza data is available only during flu season that typically starts at week 40 and ends on week 17 of the following year, we assume that the weekly flu count is zero outside the season.

2.2 Environmental Data

Meteorological and environmental parameters used in this study were obtained from two sources: satellite-derived measurements and ground stations. For satellite-derived data, we obtained daily Land Surface Temperature (LST) at 0.05° spatial resolutions from MODerate resolution Imaging Spectroradiometer (MODIS) (NASA, 2009). Both of NASA’s Terra and Aqua missions carry this instrument. We also retrieved precipitation measurements from Tropical Rain Measuring Mission (TRMM) using NASA’s Goddard Earth Sciences – Data Information Service Center (GES-DISC) Interactive Online Visualization And aNalysis Infrastructure (GIOVANNI) (Acker and Leptoukh, 2007). The TRMM data has daily temporal resolution and 0.25° spatial resolutions.

Ground station data was retrieved from Hong Kong Observatory (2009) and The Flood Control District of Maricopa County (2009). The data for Hong Kong includes air temperature (maximum, mean, minimum), mean dew point, mean relative humidity, global solar radiation and total evaporation. While for Maricopa County it includes daily mean air temperature, dew point (minimum, mean and maximum), relative humidity (minimum and maximum), maximum wind speed, air pressure (minimum and maximum) and maximum solar radiation.

In order to synchronize the temporal and spatial resolutions of the environmental data with that of influenza data, we took the spatial and 7-day average of the environmental data to form the weekly composite.

2.3 Modelling Approach

As previously mentioned, we used classical time series regression, ARIMA, to model weekly influenza cases in the two regions. This choice is due to the technique’s capability to accommodate for two of the inherent influenza count data properties, namely, seasonality and autocorrelation. In ordinary linear regression, autocorrelation behaviour generally violates the underlying assumption of the model, and hence ARIMA is more appropriate. In this section we will briefly delineate the steps in Box-Jenkins based ARIMA employed in this study. A

more detailed description can be found in (Box et al., 2009; Chatfield, 2000).

ARIMA model is based on the stationary time series assumption. That is the time series' mean and variance are independent of time. Since our data do not meet this requirement, we stabilize them by taking the logarithmic transformation and taking the difference until they are stationary. There are basically 2 terms in ARIMA model: the autoregressive (AR) and the moving average (MA) terms. The order of AR and MA are initially determined through Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Based on the ACF and PACF, we fitted several MA and AR with different orders until the residuals no longer exhibit autocorrelation behaviour. If there are several models (with different MA and AR orders) meet these criteria, we select one with the lowest Akaike's Information Criteria (AIC) or Root Mean Squared Errors (RMSE).

Since we are incorporating environmental variables as covariates, before applying ARIMA to the influenza time series, we need to assess the cross-correlation function between environmental variables and influenza. This step systematically selects the initial environmental variables – and the lags – that should be included as covariates. However, since most of the environmental variables also exhibit autocorrelation behaviour, we first modelled them using ARIMA described above, so as to ensure that the autocorrelation property does not effect the cross-correlations between the time series. For each environmental variable, we then applied ARIMA model – a step called pre-whitening – and subsequently used the residuals to calculate the cross-correlation with influenza time series. Significance was determined using two standard error limits (significant at 0.05 level). Environmental variables with significant cross-correlations were further considered in as covariates.

The next step was to apply ARIMA to influenza time series with environmental variables as the external covariates. We first included the environmental variables that express significant cross-correlations one at a time, followed by a combination of them and further selected based on the p-value (<0.05). Any models that result in residuals exhibiting autocorrelations were discarded. The models were then evaluated on the basis of their AIC and RMSE. The ARIMA model generally can be written as:

$$Z_t + \sum_{p=1}^P \phi_p Z_{t-p} = \mu + \sum_{i=1}^I \sum_{n=1}^N \beta_{i,n} X_{i,t-n} + \sum_{q=1}^Q \theta_q \varepsilon_{t-q}$$

Where, Z_t is the influenza count at week t , Z_{t-p} is the count at previous p week (AR term), μ is the intercept, $X_{i,t-n}$ is environmental variable i with lag n and ε_{t-q} is the process error at lag q (MA term). The parameters $\phi_p, \beta_{i,n}, \theta_q$ are estimated using least square method during fitting process.

An ARIMA model is notated as ARIMA(p,d,q), where p indicates the AR order, d the differencing order and q the MA order. An ARIMA model that incorporates seasonality is referred as SARIMA(p,d,q)(P,D,Q) where P,D,Q indicates the seasonal order of AR, differencing, and MA. Since the influenza time series is recorded as weekly observations, the

seasonality period is 52. All ARIMA modelling and statistical tests were performed using SAS software, Version 9.1.2 for the SAS System for Windows (SAS Institute, Inc., Cary, NC).

3. RESULTS

We first assessed the cross-correlation function of all environmental variables – both satellite-derived and ground stations measurements – with the influenza time series. We found significant correlations with LST (lag 5 wks), precipitation (lags 3, 5) and relative humidity (lags 0 to 3) for influenza in Hong Kong. Meanwhile influenza in Maricopa County was significantly correlated with LST (lag 3), mean temperature (lag 7), maximum relative humidity (lags 3, 6), minimum relative humidity (lag 6) and maximum pressure (lag 6).

Using the identified environmental variables above, we further fitted ARIMA model. Table 2 summarizes several of the model performances. These models do not result in residuals exhibiting autocorrelation behaviour.

ARIMA Model Order	Env. Vars	AIC	RMSE	
			Fit	Pred.
<i>Hong Kong</i>				
ARIMA(1,0,0)(0,1,0)	LST	134.7	0.47	0.51
ARIMA(2,1,0)	RF	168.6	0.42	0.40
ARIMA(1,0,1)(0,1,0)	RH	152.7	0.49	0.58
ARIMA(0,1,2)(1,0,0)	LST,RF,RH	137.5	0.36	0.54
<i>Maricopa County</i>				
ARIMA(2,0,0)(0,1,0)	T _{mean}	170.9	0.55	0.60
ARIMA(1,0,0)(0,1,0)	RH _{max}	207.6	0.62	0.61
ARIMA(0,1,0)(1,0,0)	P _{max}	247.4	0.54	0.51
ARIMA(1,0,1)(0,1,0)	LST	196.9	0.61	0.61
ARIMA(1,0,0)	LST, RH _{max}	277.8	0.57	0.55

Table 2 Model performances. See section 2.3 for ARIMA notation. Abbreviations: LST Land Surface Temperature; RF Rainfall (Precipitation); RH Relative Humidity; T Air temperature; P Atmospheric Pressure

As we can see from the table, the model with LST as its covariate gives the best AIC value. However, the best RMSE for the fitted dataset results from the model with LST, precipitation and relative humidity as the covariates. Since there is not much difference in the AIC values between the two models, we prefer the one with better RMSE value. That is the ARIMA model with LST, precipitation and relative humidity altogether as covariates. Figure 1A show the 95% confidence interval of the predicted and fitted values from this selected model.

For Maricopa County, the best AIC is given by ARIMA model with mean air temperature as its covariate, whereas the model with maximum pressure (P_{max}) results in the best RMSE (Table 2). However, the AIC for the model with P_{max} (AIC=247.4) is much higher than that of with T_{mean} (AIC=170.9). Thus we prefer to use the one with T_{mean} as its covariate. Figure 1B shows the resulting 95% confidence interval of the fitted and forecasted values from this model.

Using the selected model, we forecasted influenza cases 1-week ahead for the entire season. As we can see in Figure 1, the model can predict influenza peak timing reasonably well for both Hong Kong and Maricopa County.

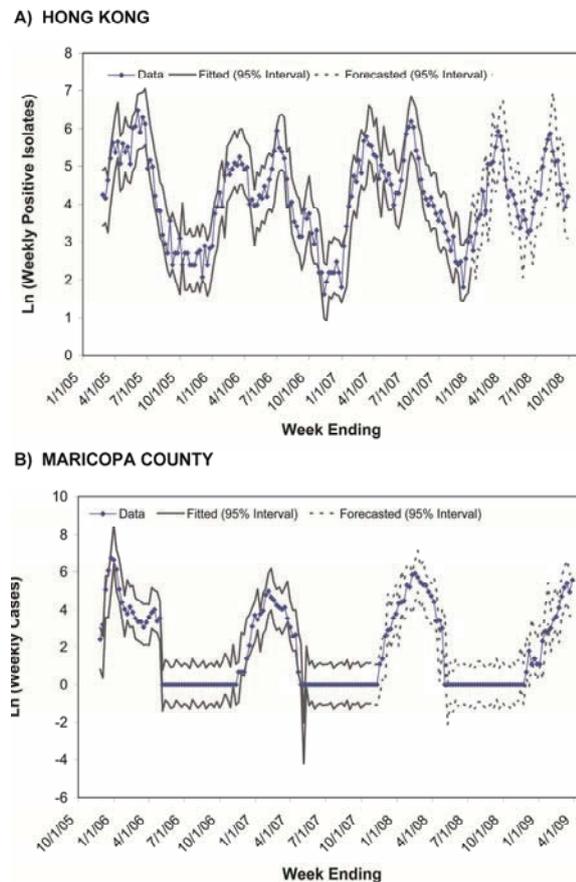


Figure 1 95% interval of fitted and predicted values

4. CONCLUSIONS

We have demonstrated here the use of environmental data in modelling influenza incidences in regions with warm climate, Hong Kong (China) and Maricopa County (Arizona, USA). We utilized ARIMA modelling approach that resulted in forecasts that are in good agreement with the actual data.

The best model for Hong Kong uses LST, precipitation and relative humidity as its covariates. These three environmental variables have been frequently implicated in influenza transmission as we previously described in section 1 and Table 1. The use of precipitation as a measure for rainfall in the selected model corroborates the role of rainfall in the tropic that has been demonstrated by other countries in the region (Chew et al., 1998; Chumkiew et al., 2007; Moura et al., 2009). For Maricopa County, we found that weekly influenza cases can be best modelled using mean air temperature as its covariates. Again, temperature has been repeatedly shown to contribute to influenza transmission. Note that Maricopa County is characterized by much less humidity than Hong Kong. Thus it is expected that rainfall or humidity do not have as much

contribution to the influenza dynamics, as reflected in the best resulting model here.

In Figure 1 we showed the 95% interval of the forecasted influenza cases using the best model selected. This result can be used by the public health organization in planning for effective prevention and control efforts such as vaccination – when and where to distribute the vaccines. The forecasts can also be used to indicate whether outbreak has occurred or has it been expected. For example, surveillance data exceeding the upper limit of the forecasts could sign an outbreak. Hence, appropriate control measures can be implemented immediately.

5. REFERENCE

- Acker, J.G. and Leptoukh, G., 2007. Online analysis enhances use of NASA Earth Science Data. *Transactions of the American Geophysical Union*, 88(2): 14-17.
- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., 2009. *Time Series Analysis: Forecasting and Control*. Wiley.
- CDC, 2010. Key facts about seasonal influenza (flu). <http://www.cdc.gov/flu/keyfacts.htm> (accessed March 2010)
- Chan, P.K.S. et al., 2009. Seasonal influenza activity in Hong Kong and its association with meteorological variations. *Journal of medical virology*, 81: 1797-1806.
- Charland, K.M. et al., 2009. Effect of environmental factors on the spatio-temporal patterns of influenza spread. *Epidemiol Infect*, 137(10): 1377-87.
- Chatfield, C., 2000. *Time Series Forecasting*. Chapman & Hall/CRC.
- Chew, F.T., Dorasingham, S., Ling, A.E., Kumarasinghe, G. and Lee, B.W., 1998. Seasonal trends of viral respiratory tract infections in the tropics. *Epidemiol Infect*, 121(1): 121-8.
- Chumkiew, S., Srisang, W., Jaroensutasinee, M. and Jaroensutasinee, K., 2007. Climatic Factors Affecting on Influenza Cases in Nakhon Si Thammarat. *Proceedings of world academy of science, engineering and technology*, 21: 364-367.
- Dowell, S.F., 2001. Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerging Infectious Diseases*, 7(3): 369-374.
- Finkelman, B.S. et al., 2007. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PLoS One*, 2(12): e1296.
- Flood Control District of Maricopa, 2009. Weather Information. <http://www.fcd.maricopa.gov/Rainfall/Weather/weather.aspx> (accessed July 2009)
- Hong Kong Department of Health, 2009. Flu Express. http://www.chp.gov.hk/en/guideline1_year/29/134/304.html (accessed January 2009)
- Hong Kong Observatory, 2009. Extract of Meteorological Observations for Hong Kong. <http://www.hko.gov.hk/wxinfo/pastwx/extract.htm> (accessed January 2009)
- Liao, C.-M., Chang, S.-Y., Chen, S.-C. and Chio, C.-P., 2009. Influenza-associated morbidity in subtropical Taiwan. *International Journal of Infectious Diseases*, 13: 589-599.

- Lofgren, E., Fefferman, N.H., Naumov, Y.N., Gorski, J. and Naumova, E.N., 2007. Influenza seasonality: underlying causes and modeling theories. *J Virol*, 81(11): 5429-36.
- Lowen, A.C., Mubareka, S., Steel, J. and Palese, P., 2007. Influenza virus transmission is dependent on relative humidity. *PLoS Pathogens*, 3(10): e151
- Lowen, A.C., Steel, J., Mubareka, S. and Palese, P., 2008. High Temperature (30 C) blocks aerosol but not contact transmission of influenza virus. *Journal of Virology*, 82(11): 5650-5652.
- Maricopa Department of Public Health, 2009. Maricopa County Weekly Influenza Summary. http://www.maricopa.gov/Public_Health/epi/flu.aspx (accessed April 2009)
- Molinari, N.A. et al., 2007. The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine*, 25(27): 5086-96.
- Moura, F.E., Perdigao, A.C. and Siqueira, M.M., 2009. Seasonality of influenza in the tropics: a distinct pattern in northeastern Brazil. *Am J Trop Med Hyg*, 81(1): 180-3.
- NASA, 2009. MODIS. <http://modis.gsfc.nasa.gov> (accessed January 2009)
- NIH/National Institute of Child Health, 2008. Why the flu virus is more infectious in cold winter temperatures *Science Daily*.
- Polozov, I.V., Bezrukov, L., Gawrisch, K. and Zimmerberg, J., 2008. Progressive ordering with decreasing temperature of the phospholipids of influenza virus. *Nature Chemical Biology*, 4(4): 248-55.
- Sagripanti, J.L. and Lytle, C.D., 2007. Inactivation of influenza virus by solar radiation. *Photochem Photobiol*, 83(5): 1278-82.
- Shaman, J. and Kohn, M., 2009. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc Natl Acad Sci U S A*, 106(9): 3243-8.
- Shaman, J., Pitzer, J., Viboud, C., Lipsitch, M. and Grenfell, B., 2009. Absolute humidity and the seasonal onset of influenza in the Continental US. *PLoS Curr. Influenza*(Dec 18): RRN1138.
- Tang, J.W., Lai, F.Y.L. and Hon, K.L.E., 2010. Incidence of common respiratory viral infections related to climate factors in hospitalized children in Hong Kong. *Epidemiol Infect*, 138: 226-235.
- Urashima, M., Shindo, N. and Okabe, N., 2003. A seasonal model to simulate influenza oscillation in Tokyo. *Jpn J Infect Dis*, 56(2): 43-7.
- Viboud, C., Alonso, W.J. and Simonsen, L., 2006. Influenza in tropical regions. *PLoS Med*, 3(4): e89.
- WHO, 2009. Influenza (Seasonal). <http://www.who.int/mediacentre/factsheets/fs211/en/> (accessed January 2010)
- Youthao, S., Jaroensutasinee, M. and Jaroensutasinee, K., 2007. Climatic factors affecting influenza cases in southern thailand. *Proceedings of world academy of science, engineering and technology*, 21: 387-391.

6. ACKNOWLEDGEMENTS

This study is supported by NASA Applied Sciences Public Health Program (<http://nasascience.nasa.gov/earth-science/applied-sciences>).