Using Earth Observations to Understand and Predict Infectious Diseases

Radina P. Soebiyanto, PhD\textsuperscript{1,2}
Richard Kiang, PhD\textsuperscript{2}

\textsuperscript{1} Universities Space Research Association, Goddard Earth Sciences Technology and Research, Columbia, Maryland
\textsuperscript{2} NASA Goddard Space Flight Center, Greenbelt, Maryland
Objective

- Characterize relationship between disease outbreaks and environmental, meteorological parameters
- Use the relationship to forecast disease outbreaks
- Disease applications:
  - Seasonal and pandemic influenza, malaria, dengue
Schematic Approach

Data Center (i.e. NASA GES DISC; NOAA NCDC; USGS Earth Explorer, etc)

Multi-temporal earth science products
Satellite-derived products i.e. TRMM3B42, MOD11C1

Ground observations

Weekly influenza epidemiological data

Mathematical Modeling
Regression model
Neural Network
Decision Tree
Agent-Based model

Model calibration and validation

Final Output: Meteorological dependency, climate-based influenza forecast

Data processing
Data continuity and integrity checking
Additional variable/indicator derivation
Spatiotemporal aggregation

Weekly meteorological time series
Meteorological Data Processing

- Epidemiological and virological surveillance data are typically aggregated
  - Spatially: district, provincial or national level
  - Temporally: weekly or monthly

- Satellite data processing
  - Projection; masking region of interest; spatial and temporal averaging; data imputation

- Ground station processing
  - Spatial and temporal averaging; data imputation

- Create lag variables
Meteorological Data Processing

Internal database of satellite data for epidemiological analysis

- Six satellite data products
- Spatial and temporal aggregation capabilities
Influenza: The Problem

Latitudinal variation of seasonal influenza epidemics
- Temperate region: distinct annual peak in winter
- Tropical region: less distinct seasonality, multiple peaks

Southward migration in Brazil
- From low population in the tropics to dense area with temperate climate

Suggest the role/influence of environmental and meteorological factors
- Several meteorological parameters has been implicated in influenza outbreaks
  - Temperate region: low temperature and humidity
  - Tropical region: rainfall in several countries
Example: Influenza In Central America

Meteorological Data

Data Source
– Tropical Rainfall Measuring Mission (TRMM): Daily resolution at 0.25° (~ 25 km)
– Global Land Data Assimilation System (GLDAS): 3-hourly resolution at 0.25° (~ 25 km)

Precipitation: TRMM
Near Surface Temperature: GLDAS
Near Surface Specific Humidity: GLDAS

Meteorological data processing

Multi-temporal (daily) precipitation rate (TRMM) from Giovanni
Logistic regression

\[ Y_{kt} \sim Bin (N_{kt}, p_{kt}) \]

\( Y_{kt} \) is the number of samples tested positive for influenza virus in location \( k \) at week \( t \); 
\( N_{kt} \) is the total samples collected/processed from location \( k \) at week \( t \); 
\( p_{kt} \) is \( Y_{kt} / N_{kt} \)

The logit of influenza positive proportion is defined as:

\[ z_{kt} = \ln \left( \frac{p_{kt}}{1 - p_{kt}} \right) \]

The full model can be written as:

\[ z_{kt} = \alpha + \sum_{j=1}^{3} \beta_{jk} x_{jk} + \sum_{l=1}^{3} \gamma_{lk} y_{lt} + \sum_{m=1}^{4} \lambda_{mk} z_{k(t-m)} + \sum_{n=1}^{3} \theta_{nk} w_{kt} \]

- Meteorological variable (i.e. temperature, humidity, rainfall)
- Co-circulating viruses (RSV, adenoviruses) as confounding factor
- Previous weeks influenza activity
- Polynomial function of week number

Regression coefficients to be estimated
Results: Estimated Coefficients

<table>
<thead>
<tr>
<th>Country and Province</th>
<th>Adjusted Odds Ratio (95% Confidence Interval)</th>
<th>Meteorological Variable Average Period</th>
<th>Prediction</th>
<th>RMSE</th>
<th>Corr. Coeff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Temperature (°C)</td>
<td>Specific Humidity (g/kg)</td>
<td>Rainfall (mm/day)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Guatemala</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central departments</td>
<td>1.01 (0.88, 1.15)</td>
<td>0.79 (0.69, 0.91)</td>
<td>1.05 (1.01, 1.09)</td>
<td>Prev. 1–3 wks ave.</td>
<td>0.08</td>
</tr>
<tr>
<td>Western departments</td>
<td>0.94 (0.80, 1.11)</td>
<td>0.72 (0.60, 0.86)</td>
<td>1.01 (0.98, 1.04)</td>
<td>Prev. 0–1 wks ave.</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>El Salvador</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West-central departments</td>
<td>0.80 (0.70, 0.91)</td>
<td>1.18 (1.07, 1.31)</td>
<td>1.00 (0.99, 1.02)</td>
<td>Prev. 1 wk ave.</td>
<td>0.06</td>
</tr>
<tr>
<td>San Miguel</td>
<td>1.28 (0.99, 1.65)</td>
<td>1.32 (1.08, 1.63)</td>
<td>0.98 (0.92, 1.05)</td>
<td>Prev. 1–2 wks ave.</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Panama</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chiriquí</td>
<td>1.30 (0.85, 2.02)</td>
<td>1.97 (1.34, 2.93)</td>
<td>0.95 (0.87, 1.04)</td>
<td>Prev. 0–3 wks ave.</td>
<td>0.11</td>
</tr>
<tr>
<td>Panama</td>
<td>1.13 (0.80, 1.61)</td>
<td>1.44 (1.08, 1.93)</td>
<td>1.10 (1.05, 1.14)</td>
<td>Prev. 1–2 wks ave.</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Bold font indicates a statistically significant variable (p-value<0.05). RMSE is the Root Mean Squared Error and Corr. Coeff is the correlation coefficient between the observation and estimated influenza positive proportion in 2013.

The models were adjusted for: potentially confounding variables (RSV, parainfluenza and adeno viruses), previous weeks’ influenza positivity, seasonality and other possible nonlinear relationships (modeled as a polynomial function, up to degree of 3, of the week number).

doi:10.1371/journal.pone.0100659.t002

**Specific humidity** was consistently associated with influenza activity in all study locations with **bimodal** relationship:

**Proportional** relationship in Guatemala and **inverse** relationship in other locations.
Results: Training and Prediction

![Graphs showing modeled training data, prediction, and observation for different regions and variables over time.](attachment:image.png)
Neural Network

Artificial intelligence method that mimic the functioning of the brain

\[ y = f \left( \sum x_i w_i \right) \]

\( f() \) can be sigmoid function, radial basis function, etc.

Input Layer

- Temperature
- Precipitation
- Humidity

Hidden layer

Output Layer

- Influenza Activity
Neural Network (NN) and ARIMA outputs for New York City and Maricopa County (AZ)

NN model shows that ~60% of influenza variability in the US regions can be accounted by meteorological factors.
Summary: Challenges

Meteorological Data and Processing
• Changes in or heterogeneity of: location, formats, algorithm, availability (data continuity)
• Storage capacity
• Data products validation

Uncovering patterns & modeling
• Choice of mathematical and statistical models
• Each model has assumptions such that results and prediction may need to be appropriately interpreted
• Parameter constraints and prediction validation
Acknowledgment

PI: Richard Kiang (NASA GSFC)

CDC
- Marc-Alain Widdowson
- Eduardo Azziz-Baumgartner
- Wilfrido Clara

El Salvador
- Oscar Rene Sorto
- Sidia Marinero

Guatemala
- Jorge Jara
- John P. McCracken
- Leticia Castillo

Panama
- Maria E Barnett de Antinori

Database developer
- Jason Lefler

This work was funded by NASA Applied Sciences – Public Health Program and CDC Influenza Division
THANK YOU
radina.soebiyanto@nasa.gov