Using Earth Observations to Understand and Predict Infectious Diseases

Radina P. Soebiyanto, PhD[1,2]
Richard Kiang, PhD[2]

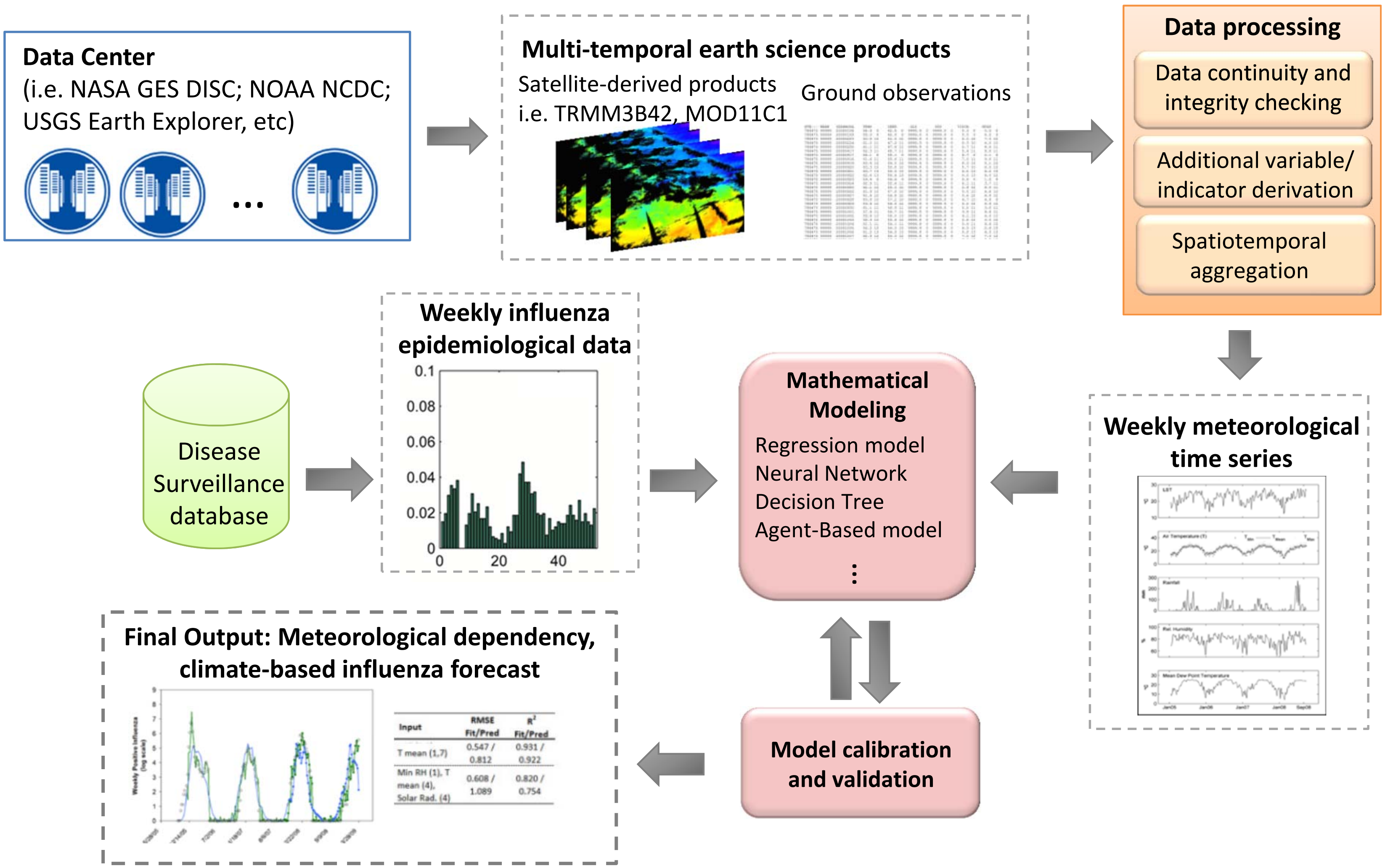
[1] Universities Space Research Association, Goddard Earth Sciences Technology and Research, Columbia, Maryland
[2] NASA Goddard Space Flight Center, Greenbelt, Maryland

ESIP Federation Meeting | January 7, 2015 | Washington DC

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



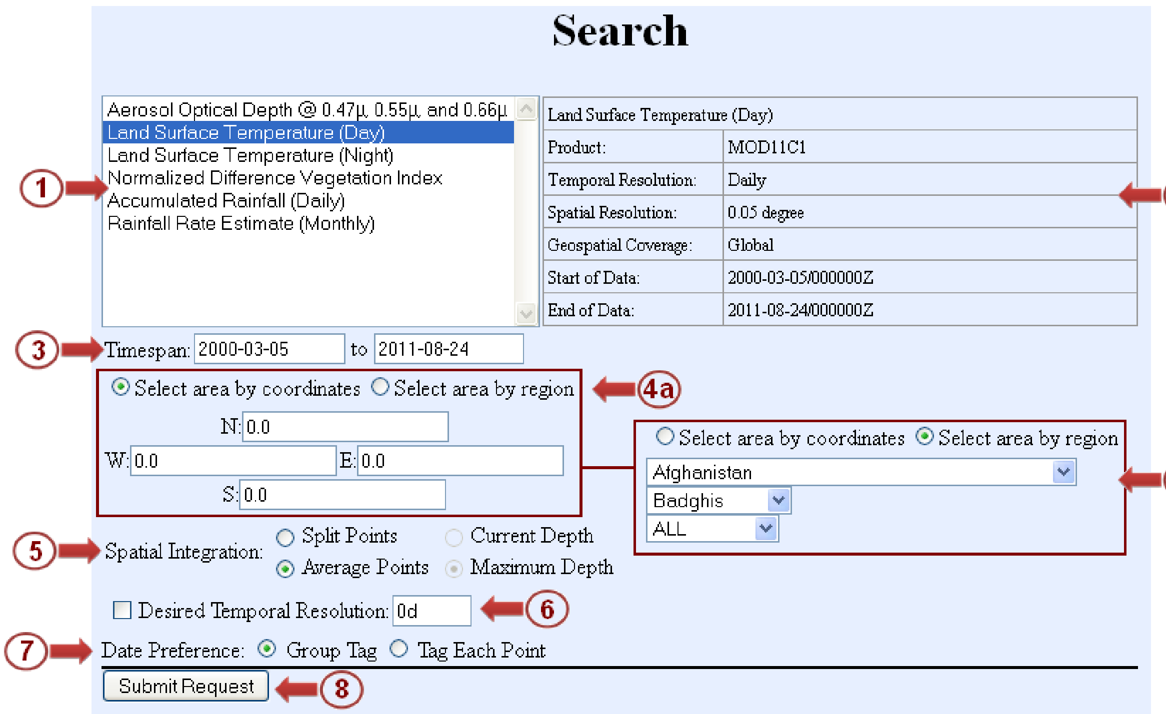
Meteorological Data Processing

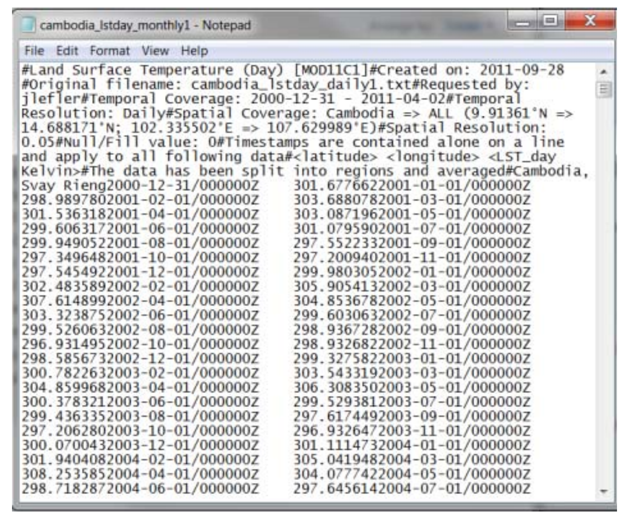
- Epidemiological and virological surveillance data are typically aggregated
  - Spatially: district, provincial or national level
  - Temporally: weekly or monthly
- Satellite data processing
  - Projection; masking region of interest; spatial and temporal averaging; data imputation
- Ground station 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





cambodia_lstday_monthly1 - Notepad

File Edit Format View Help

```
#Land Surface Temperature (Day) [MOD11C1]#Created on: 2011-09-28
#Original filename: cambodia_lstday_daily1.txt#Requested by:
jlefler#Temporal Coverage: 2000-12-31 - 2011-04-02#Temporal
Resolution: Daily#Spatial Coverage: Cambodia => ALL (9.91361'N =>
14.688171'N; 102.335502'E => 107.629989'E)#Spatial Resolution:
0.05#Null/Fill value: 0#Timestamps are contained alone on a line
and apply to all following data#<latitude> <longitude> <LST_day
Kelvin>#The data has been split into regions and averaged#Cambodia,
Svay Rieng2000-12-31/000000Z    301.6776622001-01-01/000000Z
298.9897802001-02-01/000000Z    303.6880782001-03-01/000000Z
301.5363182001-04-01/000000Z    303.0871962001-05-01/000000Z
299.6063172001-06-01/000000Z    301.0795902001-07-01/000000Z
299.9490522001-08-01/000000Z    297.5522332001-09-01/000000Z
297.3496482001-10-01/000000Z    297.2009402001-11-01/000000Z
297.5454922001-12-01/000000Z    299.9803052002-01-01/000000Z
302.4835892002-02-01/000000Z    305.9054132002-03-01/000000Z
307.6148992002-04-01/000000Z    304.8536782002-05-01/000000Z
303.3238752002-06-01/000000Z    299.6030632002-07-01/000000Z
299.5260632002-08-01/000000Z    298.9367282002-09-01/000000Z
296.9314952002-10-01/000000Z    298.9326822002-11-01/000000Z
298.5856732002-12-01/000000Z    299.3275822003-01-01/000000Z
300.7822632003-02-01/000000Z    303.5433192003-03-01/000000Z
304.8599682003-04-01/000000Z    306.3083502003-05-01/000000Z
300.3783212003-06-01/000000Z    299.5293812003-07-01/000000Z
299.4363352003-08-01/000000Z    297.6174492003-09-01/000000Z
297.2062802003-10-01/000000Z    296.9326472003-11-01/000000Z
300.0700432003-12-01/000000Z    301.1114732004-01-01/000000Z
301.9404082004-02-01/000000Z    305.0419482004-03-01/000000Z
308.2535852004-04-01/000000Z    304.0777422004-05-01/000000Z
298.7182872004-06-01/000000Z    297.6456142004-07-01/000000Z
```

Influenza: The Problem

Latitudinal variation of seasonal influenza epidemics

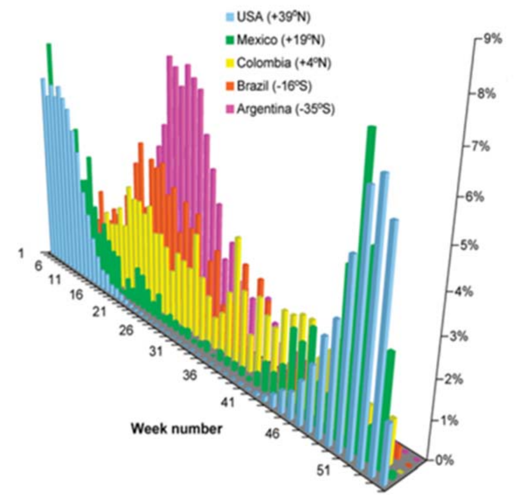
- Temperate region: distinct annual peak in winter
- Tropical region: less distinct seasonality, multiple peaks

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



Viboud et al. (2006).
PLoS Medicine 3:e89

Virus Survival	Temperature	↓
	Humidity	↓
	Solar Irradiance	↓
Transmission	Temperature	↓
	Humidity	↓
	Vapor Pressure	↓
	Rainfall	↑
	ENSO	↑
	Holidays	↑
Host Susceptibility	Sunlight	↓
	Nutrition	↕

Example: Influenza In Central America

Soebiyanto RP, Clara W, Jara J, Castillo L, et al. (2014) The Role of Temperature and Humidity on Seasonal Influenza in Tropical Areas: Guatemala, El Salvador and Panama, 2008–2013. PLoS ONE 9(6): e100659.

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)

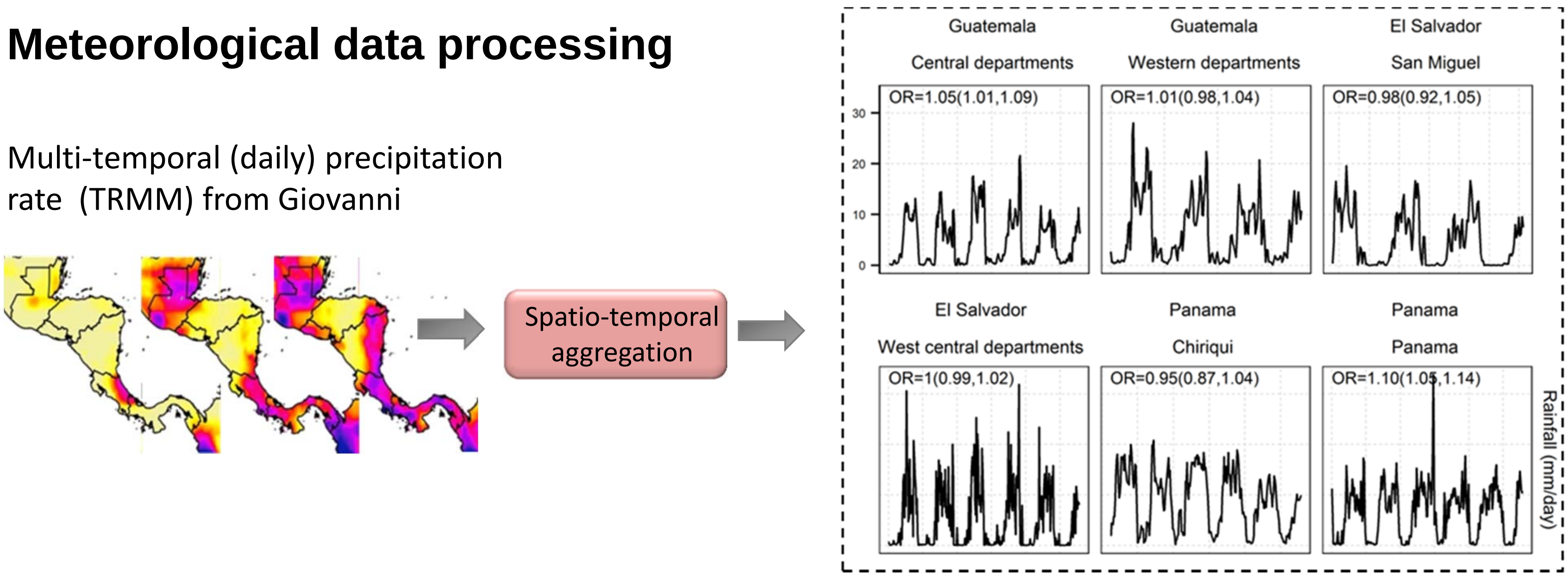
Precipitation: TRMM

Near Surface Temperature: GLDAS

Near Surface Specific Humidity: GLDAS

Meteorological data processing

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



Regression Modeling

Logistic regression

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

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

Regression coefficients to be estimated

$$z_{kt} = \alpha + \sum_{j=1}^{3} \beta_{jk} \underbrace{x_{jkt}}_{} + \sum_{l=1}^{3} \gamma_{lk} \underbrace{v_{lkt}}_{} + \sum_{m=1}^{4} \lambda_{mk} \underbrace{z_{k(t-m)}}_{} + \sum_{n=1}^{3} \theta_{nk} \underbrace{w_{kt}^{n}}_{}$$

Meteorological variable
(i.e. temperature, humidity, rainfall)

Co-circulating viruses (RSV,
adenoviruses) as confounding factor

Previous weeks
influenza activity

Polynomial function of
week number

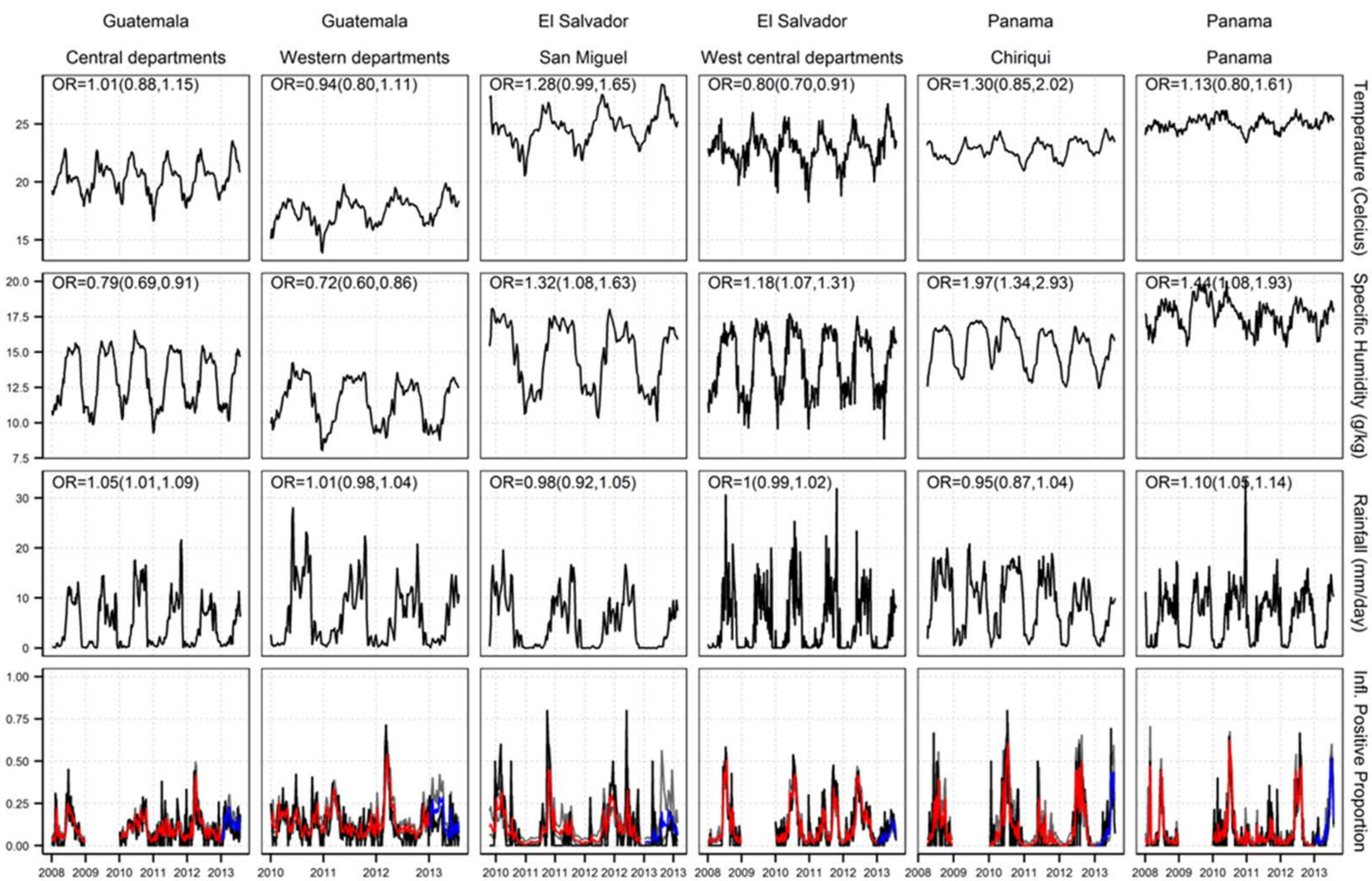
Results: Estimated Coefficients

Country and Province	Adjusted Odds Ratio (95% Confidence Interval)			Meteorological Variable Average Period	Prediction	
	Temperature (°C)	Specific Humidity (g/kg)	Rainfall (mm/day)		RMSE	Corr. Coeff
Guatemala						
Central departments	1.01 (0.88, 1.15)	**0.79 (0.69, 0.91)**	**1.05 (1.01, 1.09)**	Prev. 1–3 wks ave.	0.08	0.12
Western departments	0.94 (0.80, 1.11)	**0.72 (0.60, 0.86)**	1.01 (0.98, 1.04)	Prev. 0–1 wks ave.	0.13	0.08
El Salvador						
West-central departments	**0.80 (0.70, 0.91)**	**1.18 (1.07, 1.31)**	1.00 (0.99, 1.02)	Prev. 1 wk ave.	0.06	0.50
San Miguel	1.28 (0.99, 1.65)	**1.32 (1.08, 1.63)**	0.98 (0.92, 1.05)	Prev. 1–2 wks ave.	0.13	0.02
Panama						
Chiriquí	1.30 (0.85, 2.02)	**1.97 (1.34, 2.93)**	0.95 (0.87, 1.04)	Prev. 0–3 wks ave.	0.11	0.73
Panama	1.13 (0.80, 1.61)	**1.44 (1.08, 1.93)**	**1.10 (1.05, 1.14)**	Prev. 1–2 wks ave.	0.07	0.90

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

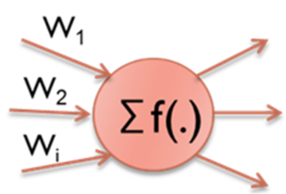


— Modeled training data — Prediction — Observation

Neural Network

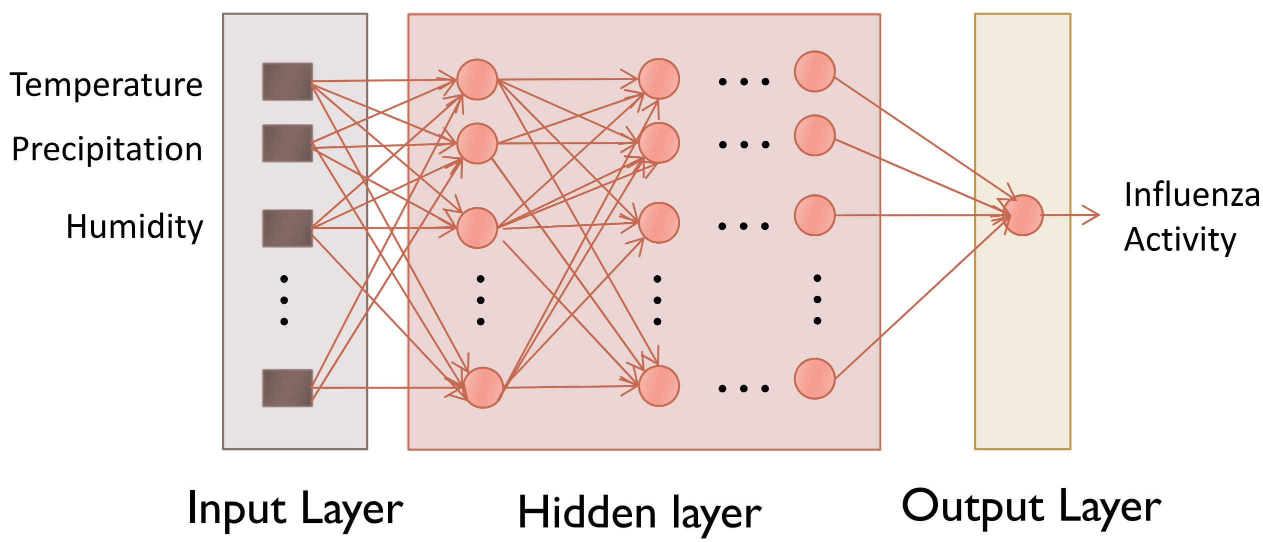
Artificial intelligence method that mimic the functioning of the brain

Artificial neuron



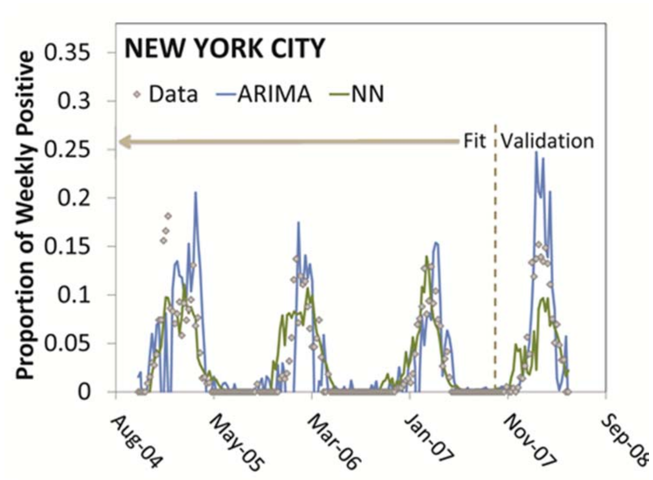
$$y = f\left(\sum x_i\, w_i\right)$$

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

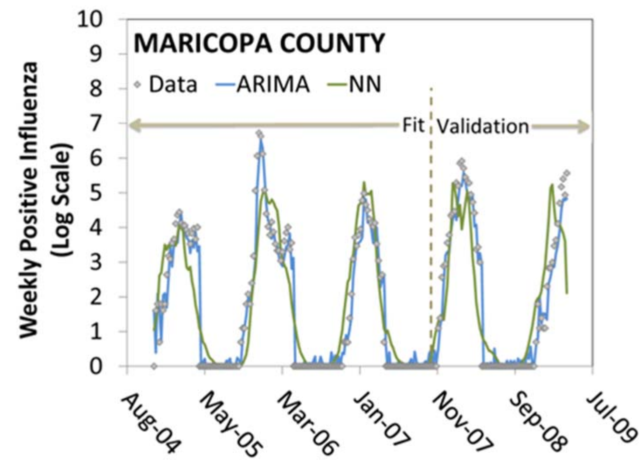


Neural Network Example

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



	Input	RMSE (Fit/Pred)	R^2 (Fit/Pred)
ARIMA	Mean Dew Pt (4)	0.046/0.022	0.311/0.795
NN	TMAX (1), Rain (3), TMIN (2)	0.044/0.0036	0.731/0.584



	Input	RMSE (Fit/Pred)	R^2 (Fit/Pred)
ARIMA	RHMAX (3), LST(3)	0.575/0.5493	0.911/0.941
NN	RHMIN(1), TEMP(4), SOLAR(4)	0.608/1.089	0.820/0.754

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

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

El Salvador
- Oscar Rene Sorto
- Sidia Marinero

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