Public Health Agency of Canada | Agence de la santé publique du Canada

Canada

ANALTYICAL TOOLS PROVIDING EPIDEMIOLOGICAL INSIGHT FOR EVENT-BASED SURVEILLANCE

2019 EIOS Global Technical Meeting
12-14 November 2019 | Seoul, Republic of Korea

PROTECTING AND EMPOWERING CANADIANS TO IMPROVE THEIR HEALTH

InSIGHT

Internet-based Surveillance Informing Global Health Threats

PRESENTATION OUTLINE

Exploiting Twitter and Google Trends
Exploiting case count data
Exploiting air traffic data

Exploiting Twitter and Google Trends data

Jean-Philippe Gilbert
PhD student
Laval University

Data analytics for aberration detection and identification of susceptible populations

Data acquisition

Twitter data from Pruss et al (2017)

GitHub

- 1st October 2015 to 31st May 2016
- 13 million of tweets
- English, Spanish, Portuguese
- Keywords “zika”, “ZIKV” and "zica”

Proprietary platforms

NEXALOGY

- Cost
- English, Korean, Russian
- GUI limited formatting/querying

Google Trends

gTrendsR package

- Free
- Volume of search terms over time & space
- Relative volume

Current work

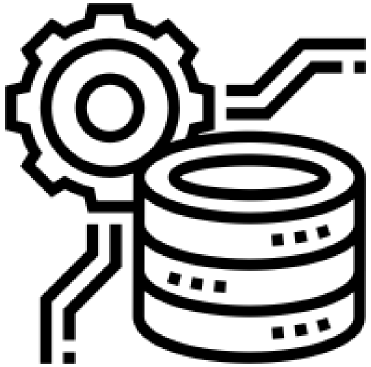
Future work

Exploiting Twitter and Google Trends data

Jean-Philippe Gilbert
PhD student
Laval University

Data analytics for aberration detection and identification of susceptible populations

Data processing



Remove re-tweets	Re-tweets by indicated by date and within Tweet	43% of remaining tweets are original
Geolocate tweets	1% users provide coordinates	50% tweets populated with coordinates (and resolution of geolocator)
Classify content of tweets	NLP and machine learning classifiers (e.g. Miller et al. 2017)	Symptoms Infection Prevention Treatment

python™
Carmen library

Current work | Future work

Exploiting Twitter and Google Trends data

Jean-Philippe Gilbert
PhD student
Laval University

Data analytics for aberration detection and identification of susceptible populations

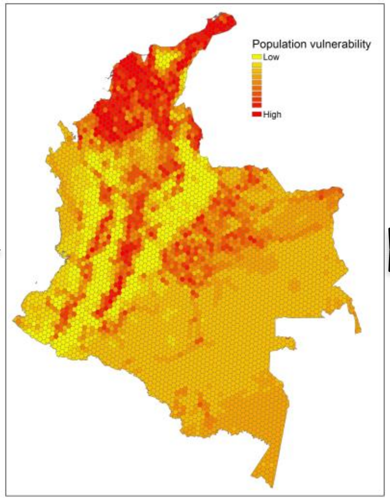
Analytical tools

Aberration detection of symptoms

- Exploit clues from EBS data
- Output from classifier

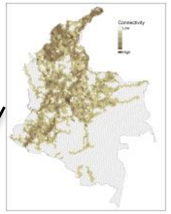
Augmenting identification of susceptible populations

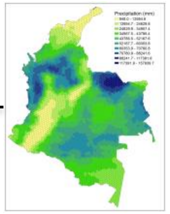
Geolocated classified tweets

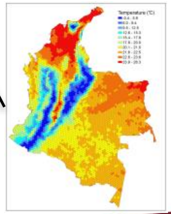


Disease susceptibility

Environmental / social determinates







Current work

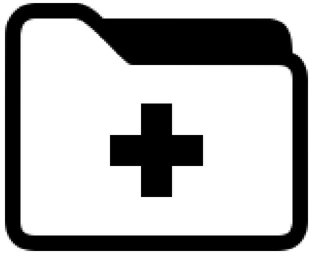
Future work

Exploiting EBS extracted case count data

Julie Simon, PhD
PHAC

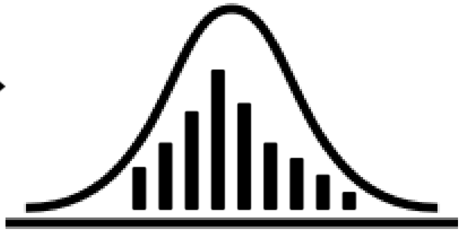
Forecasting cases through time

Case data



NLP extracted cases
Official surveillance reports

Incidence Decay and Exponential Adjustment (IDEA) model



$$I_t = \left(\left(R_0/(1+d)^t\right)^t\right.$$

Number of confirmed/suspected cases, I_t
Basic reproductive number, *R0*
Discount factor, *d*
Serial interval, *t*

- Forecasts epidemiological curve
- Estimates potential for disease spread
- Can signal change in outbreak
- Simple
- Not validated for real-time outbreaks

How few weeks of case data can provide reasonable model output?

Exploiting EBS extracted case count data

Julie Simon, PhD
PHAC

How many weeks of data are needed for reliable model output?

Simulated data
Fisman et al (2013)

SIR model output: cumulative incidence

Four serial intervals, *t*

PAHO surveillance data for ZIKV

46 countries
Jan. 2015 to Dec. 2016
Backfilled data to index case

Observed surveillance data: Cumulative incidence and frequency of cases

5-7 serial intervals, *t*;
High variation

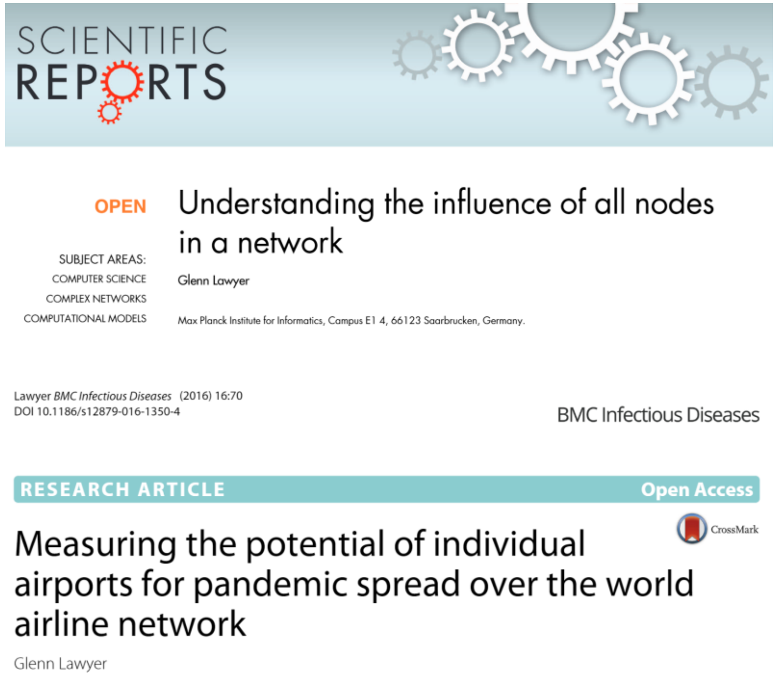
Depends on country, data transformation

Smooth case data? (e.g. Majumder et al.)

Exploiting air traffic data

Julie Simon, PhD
PHAC

- Disease spread through air travel
- Meta-population compartmental models
- Simpler tools?
- Just passenger volume data → inadequate



SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
COMPUTER SCIENCE
COMPLEX NETWORKS
COMPUTATIONAL MODELS

Understanding the influence of all nodes in a network

Glenn Lawyer

Max Planck Institute for Informatics, Campus E1 4, 66123 Saarbrucken, Germany.

Lawyer *BMC Infectious Diseases* (2016) 16:70
DOI 10.1186/s12879-016-1350-4

BMC Infectious Diseases

RESEARCH ARTICLE Open Access

CrossMark

Measuring the potential of individual airports for pandemic spread over the world airline network

Glenn Lawyer

- Passenger volume data AND network topology → better!
- Airport expected force of infection (ExF)

Can ExF predict the risk of disease spread earlier than the IDEA model?

Current work

Future work

Exploiting air traffic data

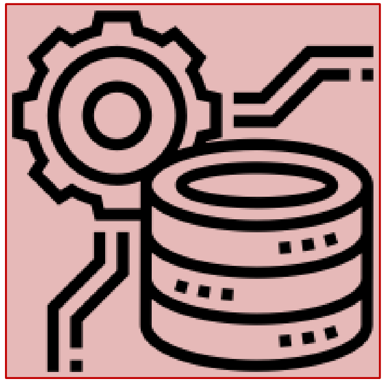
Julie Simon, PhD
PHAC

Assessing ExF as early predictor for reported travel-acquired ZIKV cases in Canada

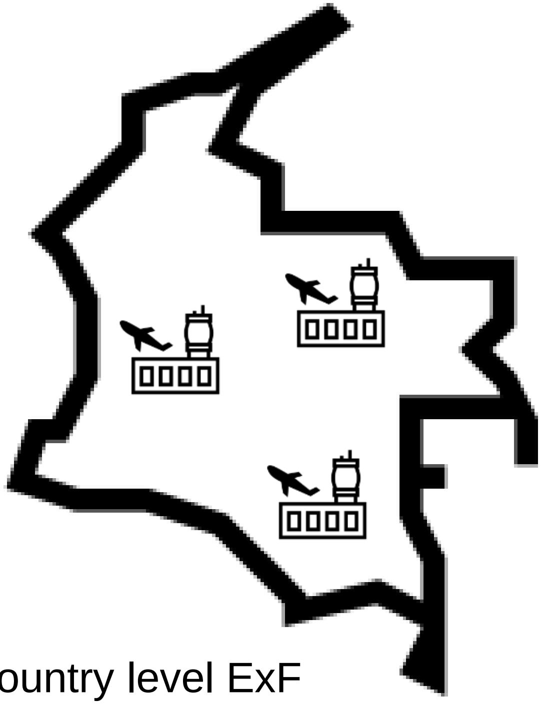
Data acquisition

- 46 countries in Americas
- Air traffic data (BlueDot/IATA)
- Reported travel-acquired cases
- Other covariates

Data processing



- Currently 29 countries
- Aggregating ExF to spatial resolution of case data



Country level ExF

- Mean ExF of airports
- Max ExF of airports
- CoV ExF of airports

Current work

Future work

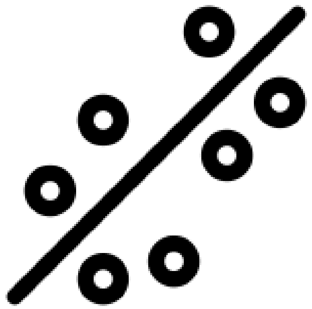
Exploiting air traffic data

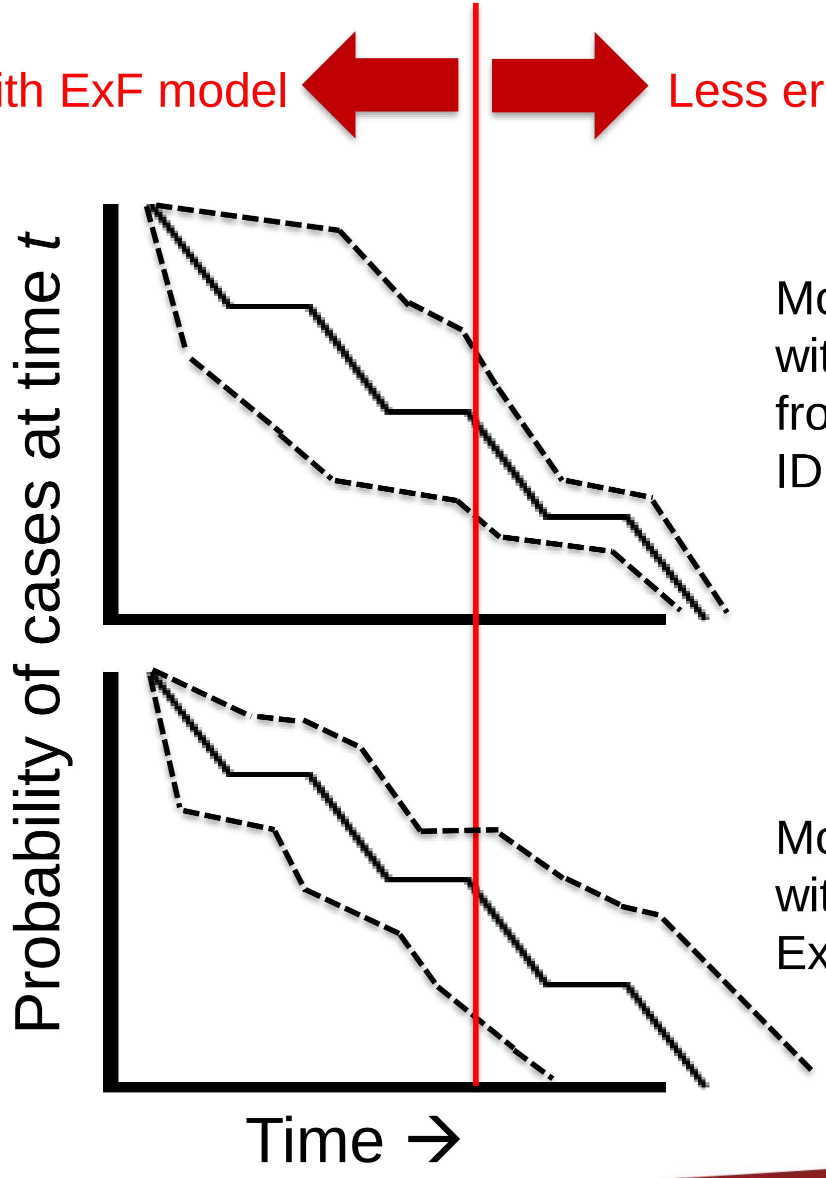
Julie Simon, PhD
PHAC

Less error with ExF model

Less error with IDEA model

Statistical modelling





Model with R0 from IDEA

Model with ExF

Current work

Future work

Closing remarks

Challenge of developing data-driven models when data are scarce

- Get InSIGHT earlier: SIR → IDEA → ExF
- Exploit data extracted from EBS systems
- Exploit other data sources
- Integrate data types where possible

InSIGHT
Internet-based Surveillance Informing Global Health Threats

Public Health Agency of Canada / Agence de la santé publique du Canada

EIOS epidemic intelligence from open sources

UQÀM Université du Québec à Montréal

Natural Resources Canada / Ressources naturelles Canada

Canada

University of Guelph — Changing Lives Improving Life

Health Canada / Santé Canada

Computer Science University of Toronto

Université Laval

McGill Department of Epidemiology, Biostatistics and Occupational Health