Predictive models for infectious diseases spread

BDVA - Are we using data in the best way to manage the COVID-19 Pandemic?

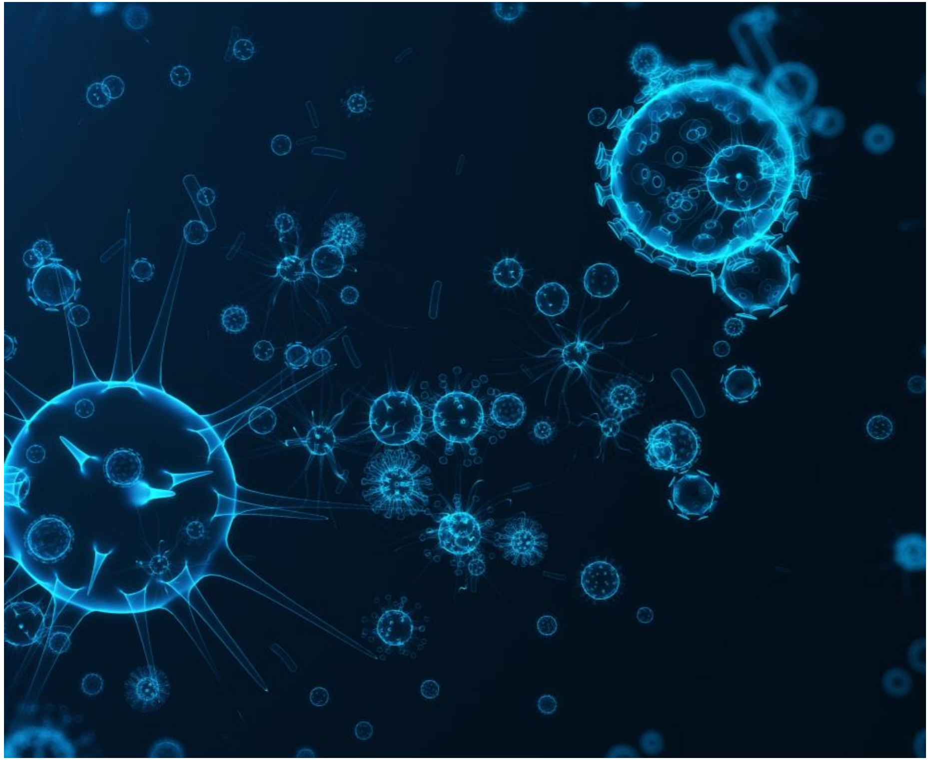
26-03-2020

Atos

Motivation

Predictive modelling for infectious diseases spread

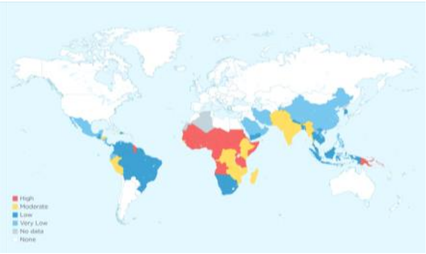
- Preventive policies
- Decision-making
- Allocation of medical resources
- Raise awareness
- Arrangement of production activities
- Economical impact
- Seasonal prediction



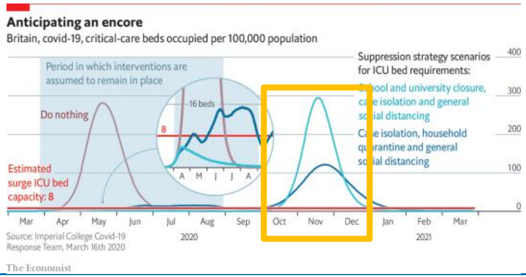
Disease prediction models

Corley CD, Pullum LL, Hartley DM, et al. Disease prediction models and operational readiness. *PLoS One*. 2014;9(3):e91989. Published 2014 Mar 19. doi:10.1371/journal.pone.0091989

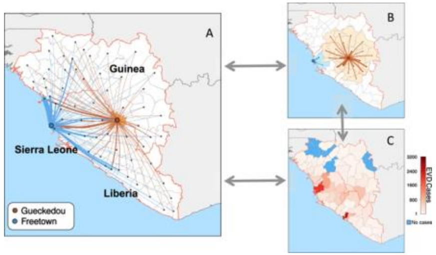
Risk assessment

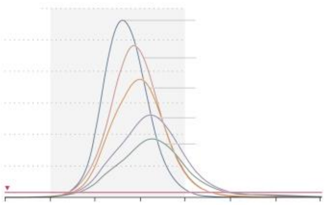


Event prediction



Spatial / dynamical models

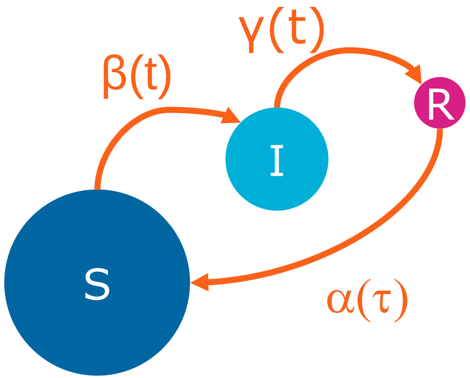


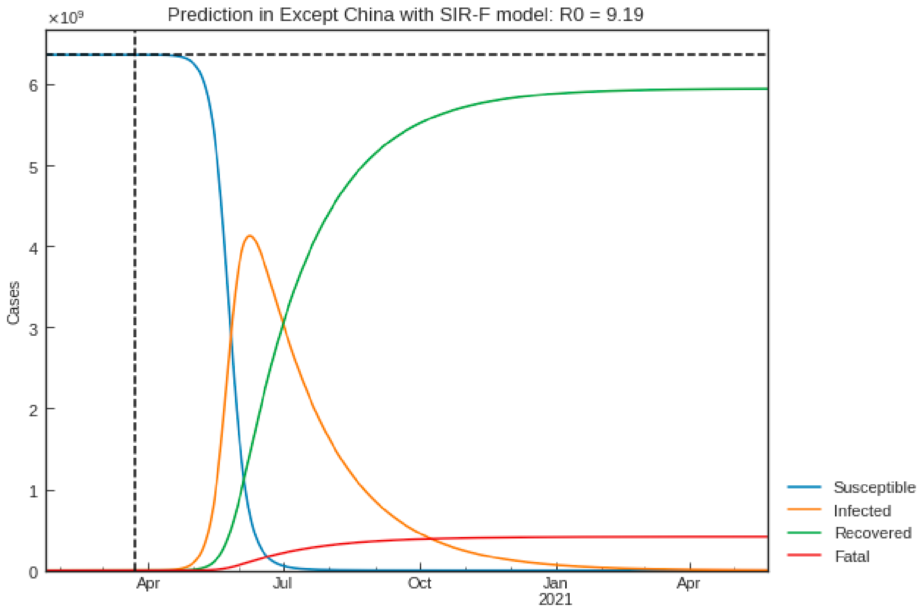


Event detection

Mathematical modelling

- SIR model (Susceptible, Infectious, Recovered)
  - Compartment model
  - Human to human transmission
  - Based on differential equations
  - Improvements and elaborations
    - SIS, carriers, SEIR (exposed), etc





Mathematical modelling

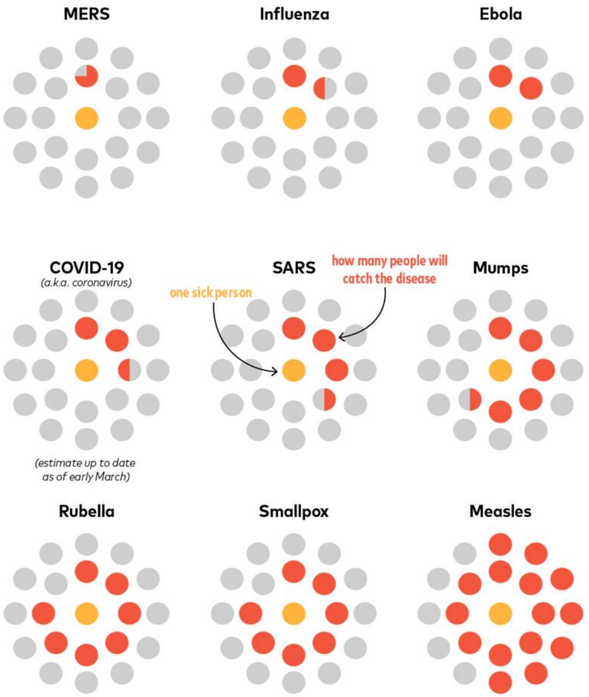
- Reproduction number (R0)

Duration of contagiousness

Likelihood of infection per contact

Contact rate

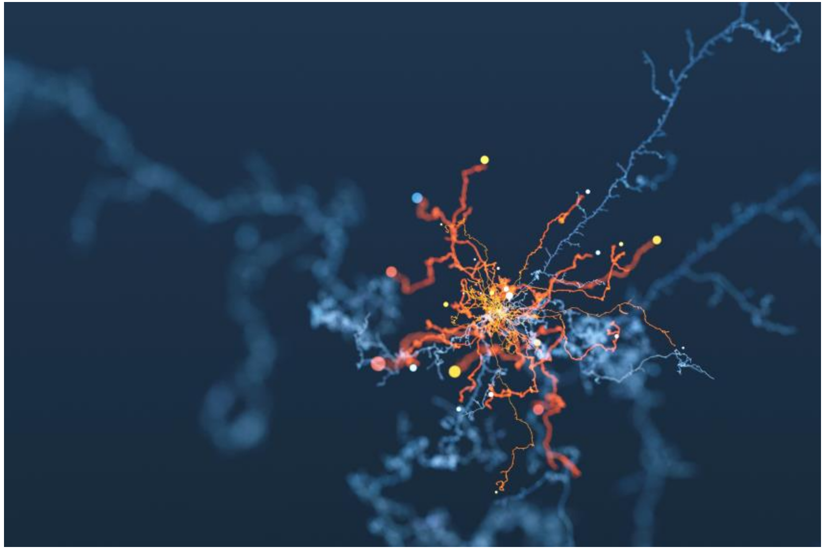
Reproduction number



https://www.popsci.com/story/health/how-diseases-spread/

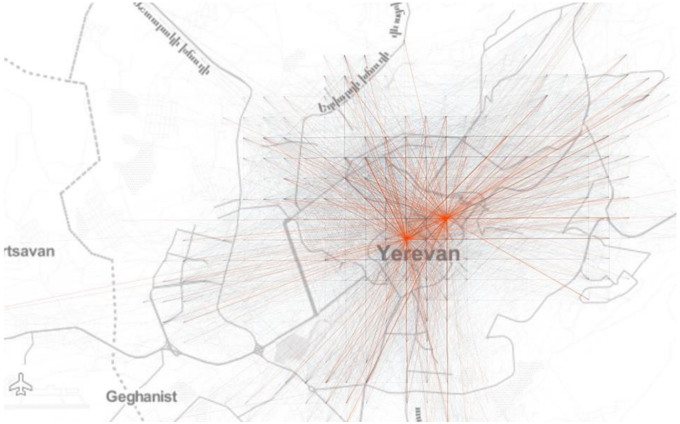
AI for spread prediction

- SIR parameters estimation
  - Machine Learning (MLP, CNN, LSTM) can be used to estimate SIR parameters from time series data as a supervised problem.
  - ML models learn faster than Approximate Bayesian Computation (ABC).
  - Tessmer, H. L., Ito, K., & Omori, R. (2018). Can machines learn respiratory virus epidemiology?: A comparative study of likelihood-free methods for the estimation of epidemiological dynamics.
  - Maziar Raissi, Niloofar Ramezani & Padmanabhan Seshaiyer (2019) On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods



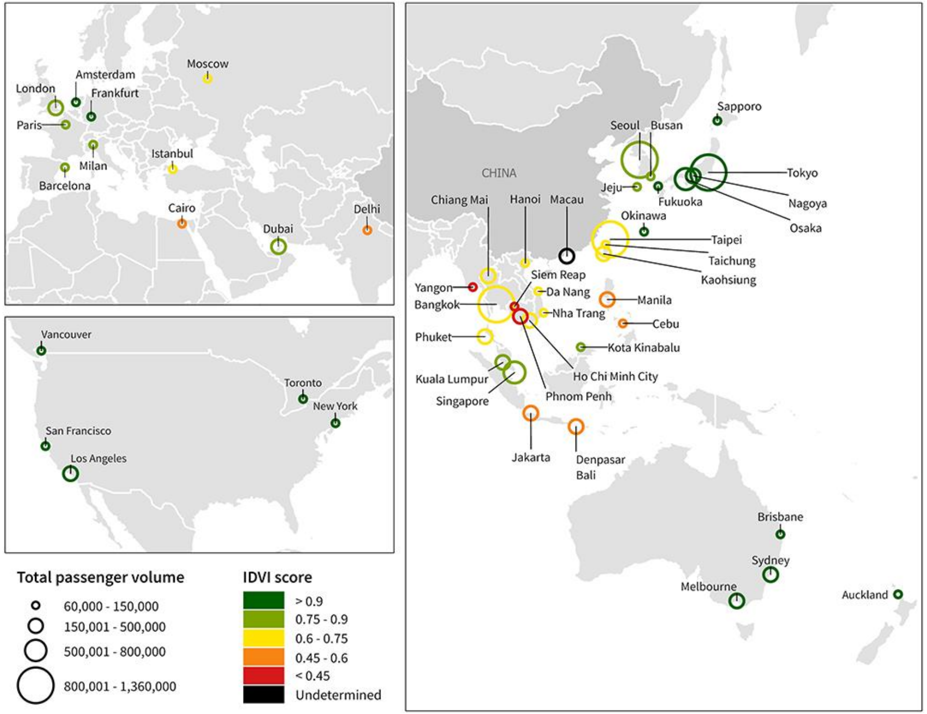
Big Data for spread prediction

- Spatiotemporal spread prediction using data
  - Traffic Origin Destination (OD) matrix can be used to predict spread considering population flows.
  - OD matrix can be obtained from real data.
  - Models show that there is a positive correlation between the population input and the number of confirmed cases.
  - https://lexparsimon.github.io/coronavirus/



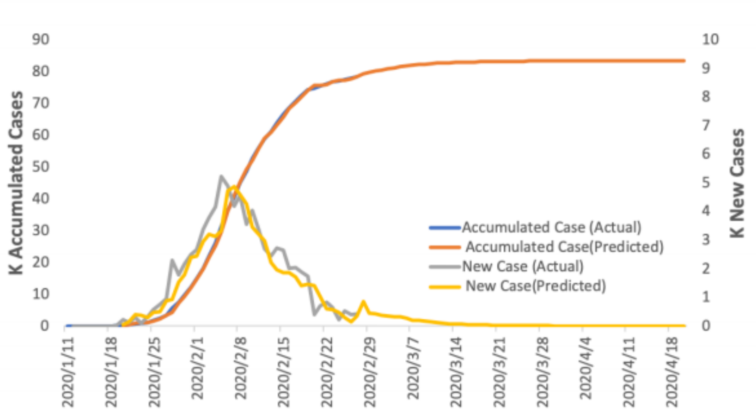
AI & BigData for spread prediction

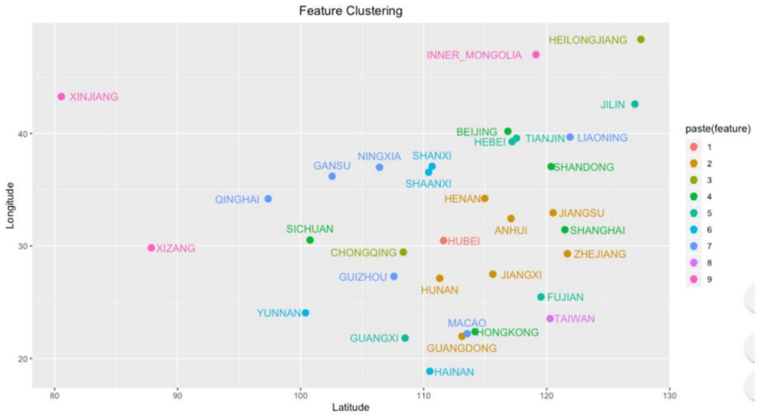
- BlueDot spread prediction
  - Published in 2020/01/27
  - Using 2019 data from the International Air Transport Association (IATA)
  - Infectious disease vulnerability index (IDVI) for each receiving country
  - 11 of the cities at the top of their list were the first places to see COVID-19 cases.
  - Journal of Travel Medicine, Volume 27, Issue 2, March 2020, taaa011



AI for spread prediction

- Real-time forecasting using auto-encoders
  - Modified auto-encoders forecast number of accumulative and new confirmed cases.
  - Clustering for provinces and cities is done using k-means.
  - Hu, Z., Ge, Q., Jin, L., & Xiong, M. (2020). Artificial intelligence forecasting of covid-19 in China.





COVID-19 datasets (global)

Dataset	Frequency	Information	Level of detail	Format
John Hopkins University	Daily	Confirmed, deaths, recovered, active	Per country and region (US)	Daily reports and Time-series (csv)
WHO situation reports	Daily	Confirmed, deaths, transmission classification	Per country	Situation report (pdf) and API
European Centre for Disease Prevention and Control	Daily	New cases and new deaths	Per country	Situation report (pdf) and Time-series (csv)

COVID-19 datasets (national)

Dataset	Frequency	Information	Level of detail	Format
Spain	Daily	Cases, hospitalized, ICU, deaths	By province	CSV
South Korea	Daily	Cases, patients info, age, gender, etc	By case	CSV
Italy	Daily	Positive cases	By province & region	CSV
Brazil	Daily	Suspects, refuses, cases, deaths	By state	CSV
Indonesia	Daily	Tested, confirmed, negative, isolated, reléased, patient metadata	By case	CSV
India	Daily	Confirmed, cured, deaths, patient metadata	By case	CSV

Challenges

Access to quality, accurate, and complete data

Small amount of available data

Effects of quarantine and actions

Variability over the time

Variability between regions

Compliance with privacy regulations

Atos

Challenges

“Right now the quality of the data is so uncertain that we don't know how good the models are going to be in projecting this kind of outbreak”

Marc Lipsitch, epidemiologist at the Harvard T.H. Chan School of Public Health

Atos

Thank you

For more information please contact:
daniel.calvo@atos.net
adrian.arroyo@atos.net