NAU NORTHERN ARIZONA UNIVERSITY

School of Informatics, Computing, and Cyber Systems

Dynamical modeling of infectious disease spread using data-driven models

Kayode Oshinubi, PhD

Post-doctoral researcher

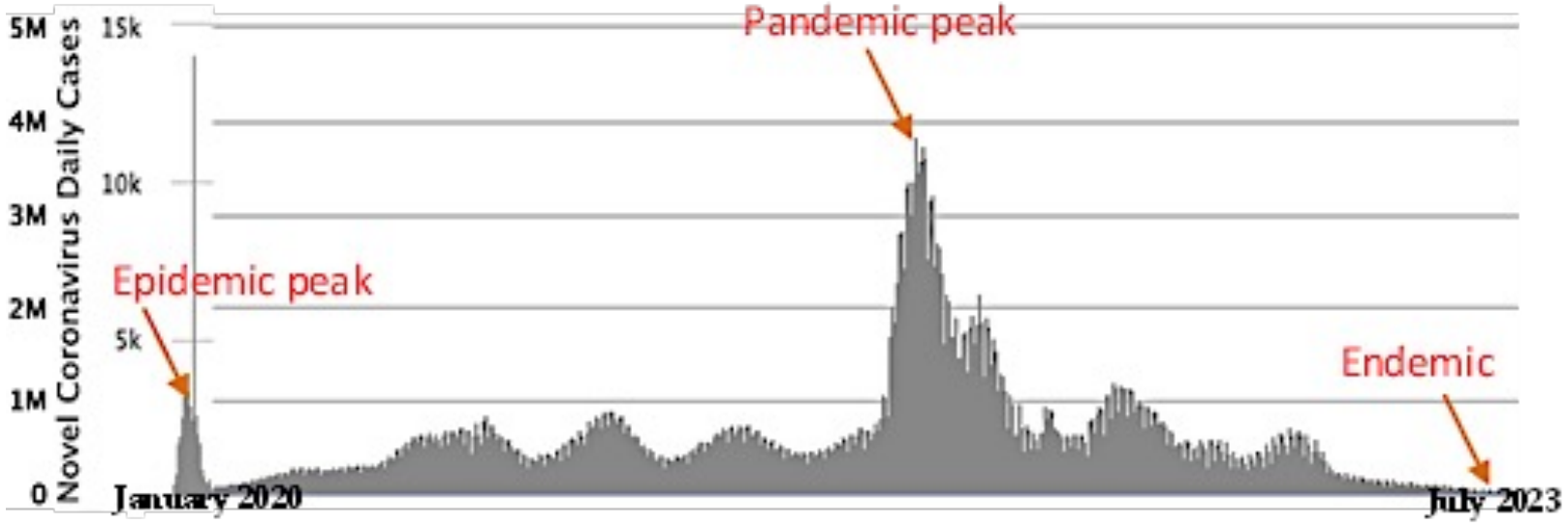
05-29-2024

Outline

- Motivation
- Background
- Our Models
- Results
- Ongoing/Future work
- Summary and conclusion
- Some publications

The motivation behind the problem: Infectious disease curve

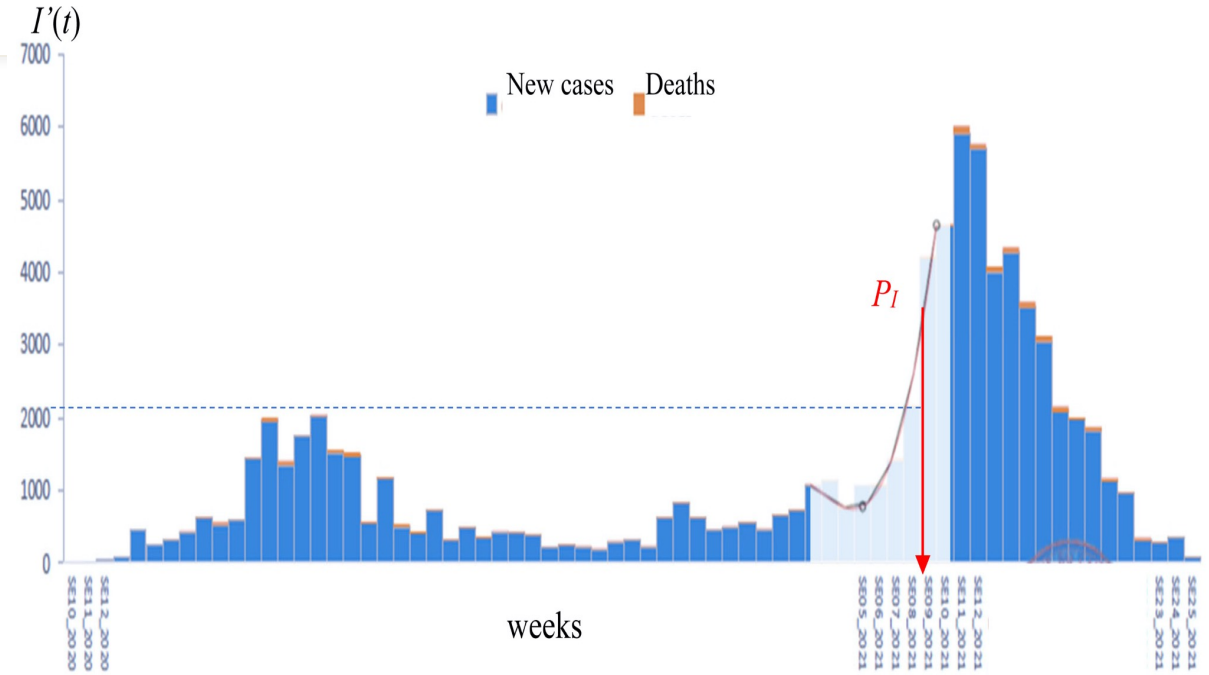
- There is a global pandemic



The pandemic outlook as at July 2023

- There are three stages of infectious disease

The motivation behind the problem: Infectious disease curve



COVID-19 weekly new cases in Cameroon showing the inflection point P_I (red arrow) during the first half part of the second wave between week 5 of 2021 and week 12 of 2021

The proposed solution

To develop a robust method for predicting the changes and transition between endemic and epidemic phases of an infectious disease, using COVID outbreak as an example

- ✓ We define indicators for detecting changes and transitions between endemic and epidemic phases using seven scalars calculated from daily reported news cases and deaths. The indicators chosen are related to the form of the empirical distribution of new cases seen over a fourteen day period chosen to smooth out the influence of weekends when fewer new cases are registered.
- ✓ We used the Principal Component Analysis to create a score from the seven proposed indicators that allows an acceptable level of forecasting performance by providing a realistic retro predicted date for the rupture of the stationary endemic model corresponding to the entry into the epidemic exponential growth phase.

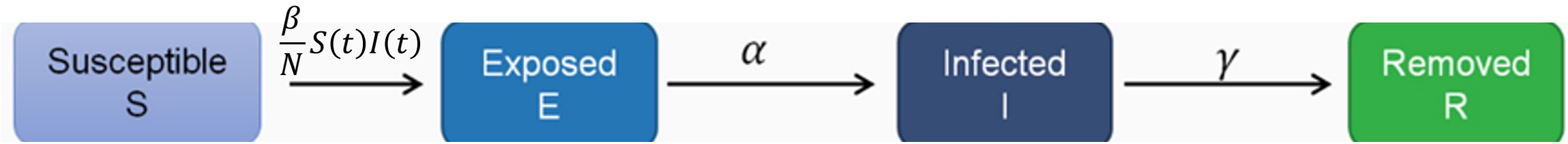
Background

The modeling of infectious disease started as early as 1760 **by Daniel Bernoulli who proposed the famous deterministic mathematical model Susceptible Infected SI to solve the epidemic wave but not modeled the endemic state of smallpox given as:**

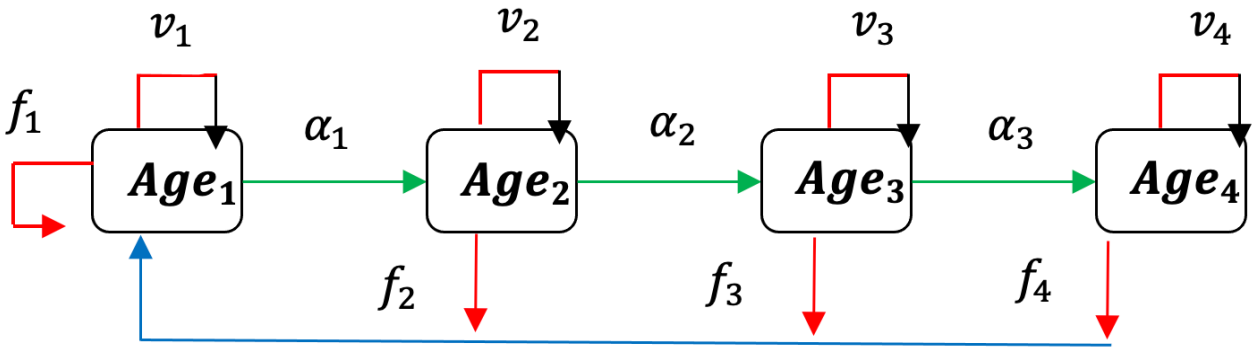
$$\frac{dS}{dt} = -\frac{\beta}{N}S(t)I(t), \qquad \frac{dI}{dt} = \frac{\beta}{N}S(t)I(t) - \nu I(t),$$

where ν is the specific mortality rate due to the disease, β/N the disease transmission rate, $S(t)$ the number of susceptible individuals, $I(t)$ the number of infectious individuals at time $t \geq 0$.

Background

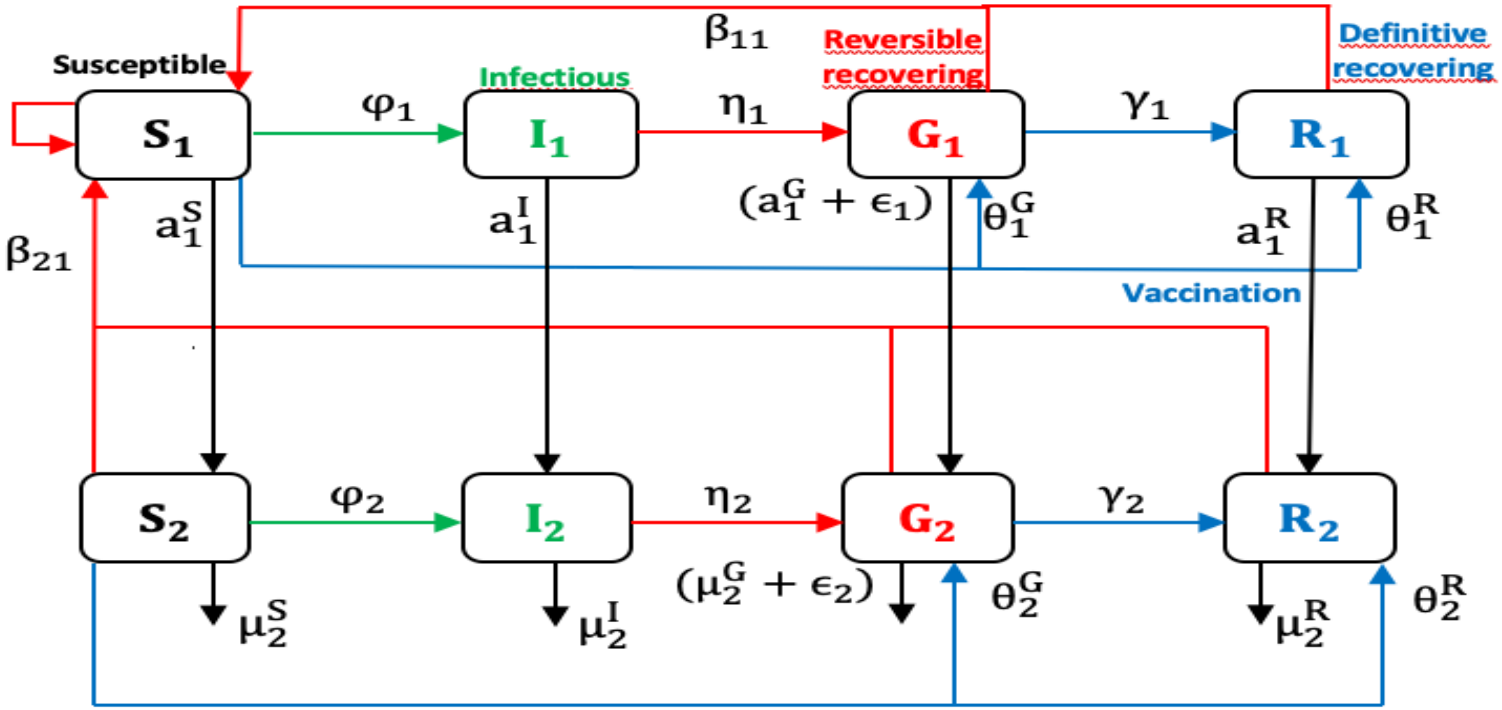


Schematic diagram of a simple SEIR model

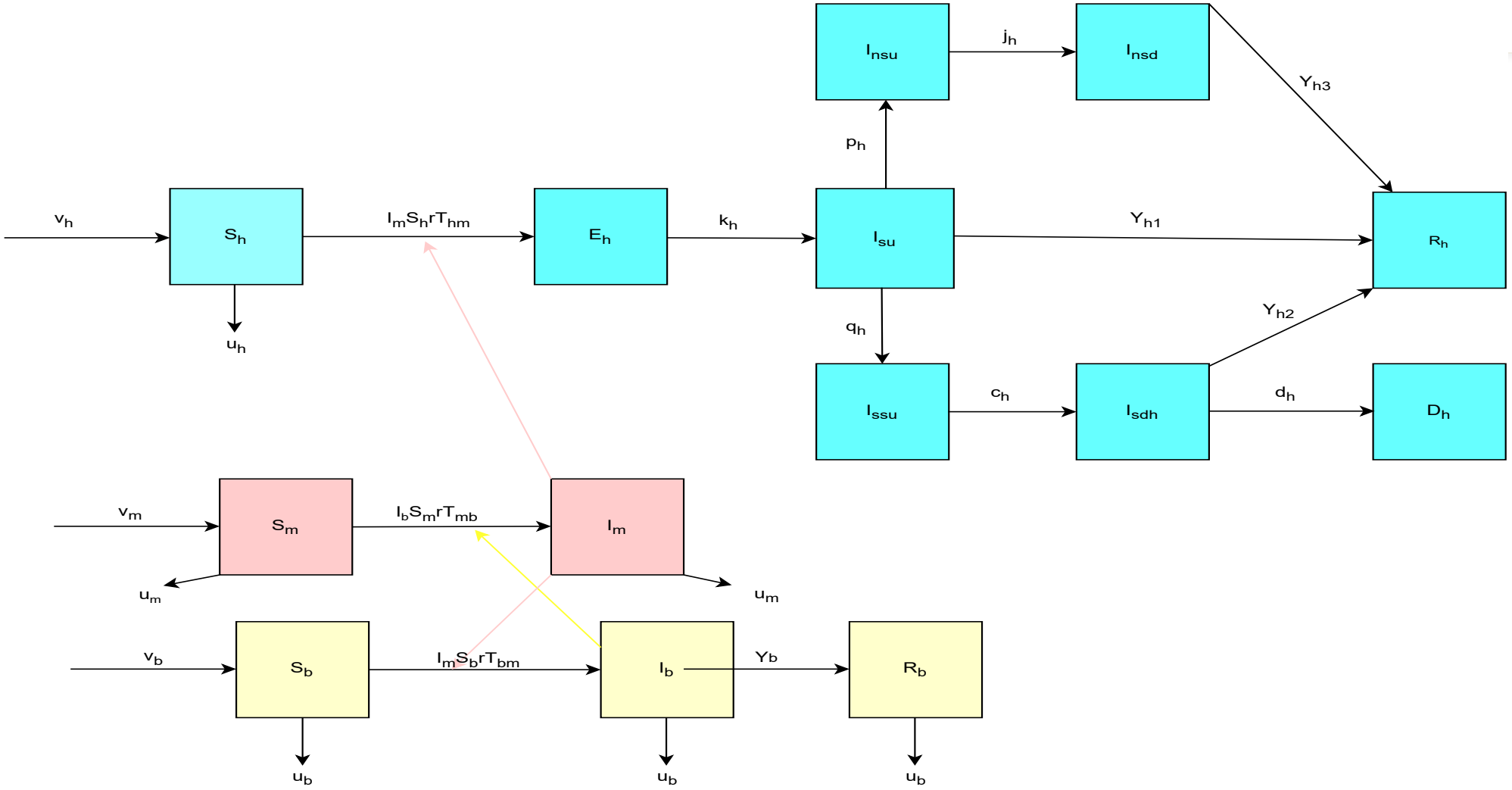


Flow chart illustrating the epidemiology ageing model with four age classes, f_1, f_2 and f_3 are non-zero fertility rates, when a disease like an epidemic outbreak concerning all the age classes occurs. ν_j is the probability of remaining in the same state and $\propto_j$ is the probability of going to another state.

Background


Susceptible
Infectious
Reversible recovering
Definitive recovering
Vaccination
S_1
I_1
G_1
R_1
S_2
I_2
G_2
R_2
β_{11}
φ_1
η_1
γ_1
β_{21}
a_1^S
a_1^I
$(a_1^G + \epsilon_1)$
θ_1^G
a_1^R
θ_1^R
φ_2
η_2
γ_2
μ_2^S
μ_2^I
$(\mu_2^G + \epsilon_2)$
θ_2^G
μ_2^R
θ_2^R

Background


I_{nsu}
j_h
I_{nsd}
Y_{h3}
p_h
v_h
S_h
$I_mS_hrT_{hm}$
E_h
k_h
I_{su}
Y_{h1}
R_h
u_h
q_h
Y_{h2}
I_{ssu}
c_h
I_{sdh}
d_h
D_h
v_m
S_m
$I_bS_mrT_{mb}$
I_m
u_m
u_m
v_b
S_b
$I_mS_brT_{bm}$
I_b
Y_b
R_b
u_b
u_b
u_b

Background: Phenomenological models

Endemic phase: During the endemic phase, the dynamics of new cases appears to fluctuate around an average value independently of the number of cases. Therefore the average cumulative number of cases is given by

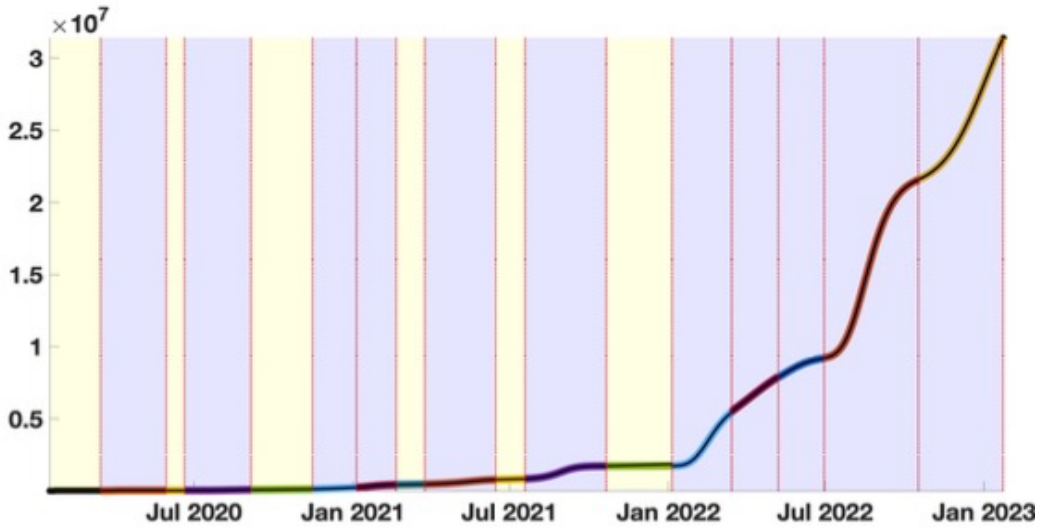
$$\mathrm{CR}(t) = N_0 + (t - t_0) \times a, \text{ for } t \in [t_0, t_1], \tag{2.8}$$

where t_0 denotes the beginning of the endemic phase, N_0 is the number of new cases at time t_0, and a is the average value of the daily number of new cases.

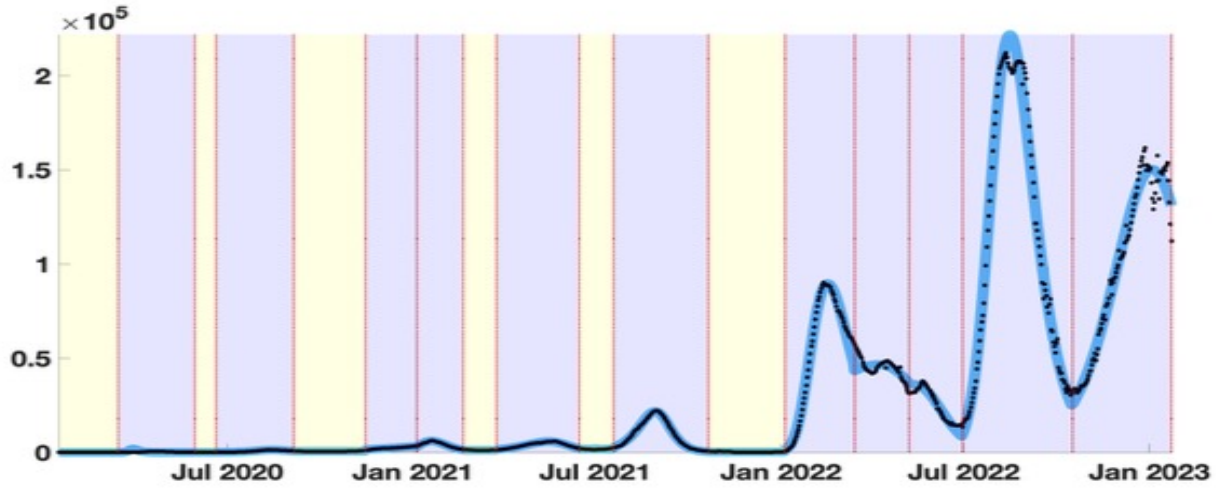
We assume that the average daily number of new cases is constant. Therefore the daily number of new cases is given by

$$\mathrm{CR}'(t) = a. \tag{2.9}$$

Background: Phenomenological models

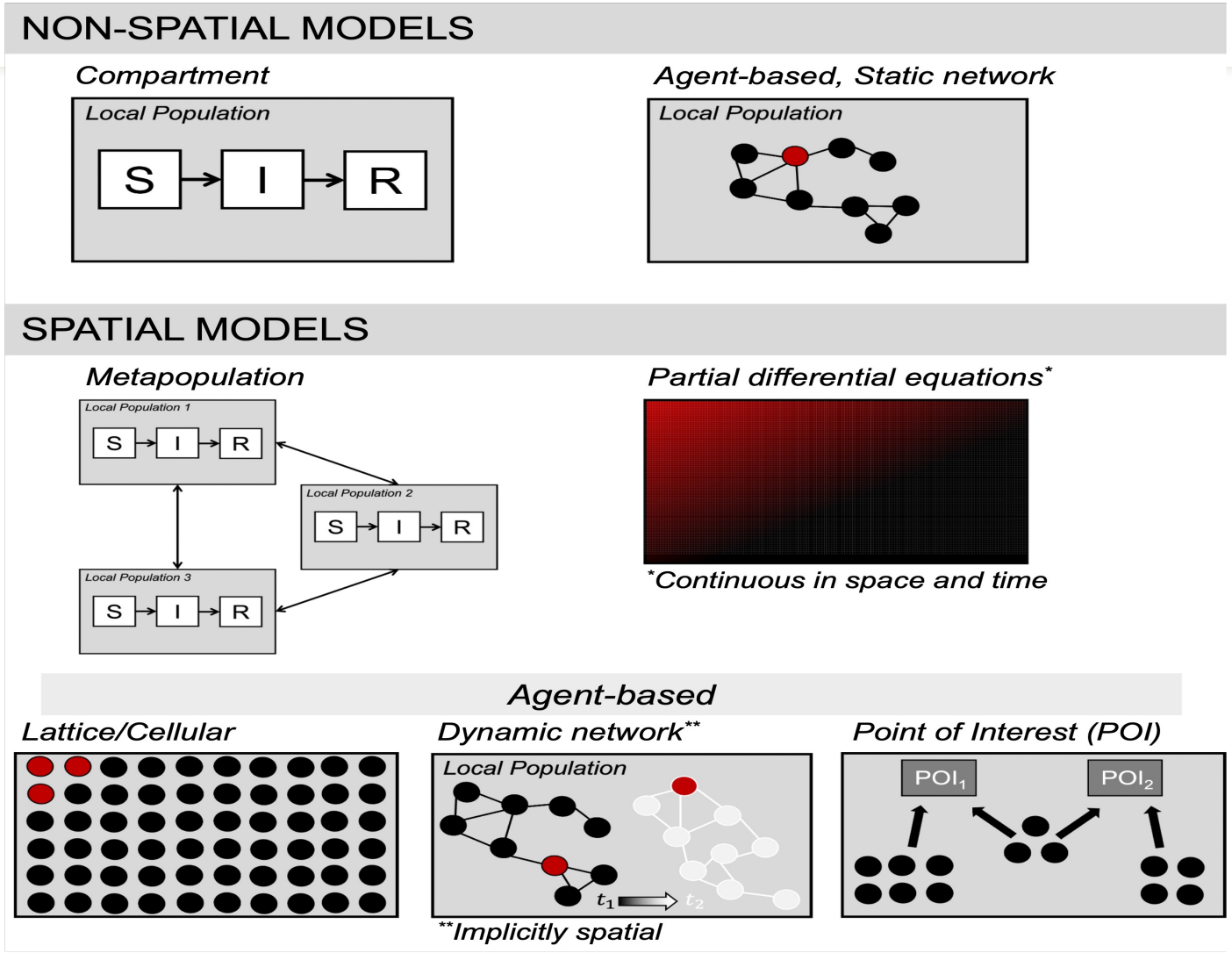


In this figure we plot with multiple colors the phenomenological models obtained for each period.



In this figure we plot in blue the first derivative of the phenomenological model and in black the data. Data is the daily reported number of new cases with a 14-day rolling average.

Background


NON-SPATIAL MODELS
Compartment
Local Population
S
I
R
Agent-based, Static network
Local Population
SPATIAL MODELS
Metapopulation
Local Population 1
S
I
R
Local Population 2
S
I
R
Local Population 3
S
I
R
Partial differential equations*
*Continuous in space and time
Agent-based
Lattice/Cellular
Dynamic network**
Local Population
t_1
t_2
Point of Interest (POI)
POI_1
POI_2
**Implicitly spatial

Models

We use in the following a moving window of length 14 days for calculating the empirical distribution of the random variable equal to the number of daily reported new cases. The empirical distribution N_t on day t is obtained from the daily number of reported new cases considered as a random variable $N_t = (N(t-13), N(t-12), \ldots, N(t))$.

We consider the first four moments of N_t. We start with the *mean*

$$\mu = E(N_t) = \frac{\sum_{i=0}^{13} N(t-i)}{14},$$

where E is the expectation operator, with the *standard deviation*

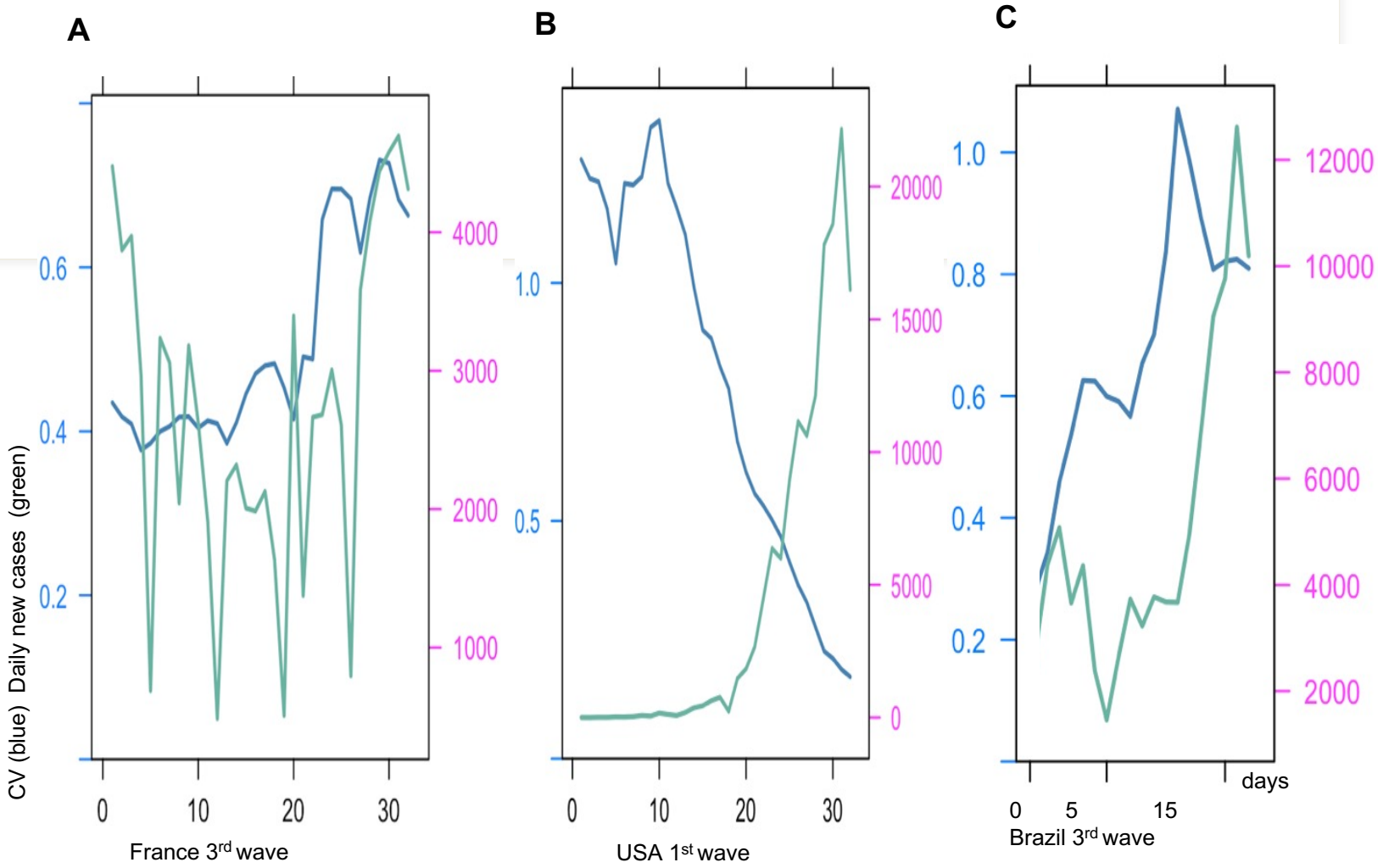
$$\sigma = E\left((N_t - \mu)^2\right)^{1/2} = \sqrt{\frac{\sum_{i=0}^{13} (N(t-i) - \mu)^2}{14}}.$$

Models

From these two first parameters, we can compute the coefficient of variation

$$CV(N_t) = \sigma/\mu$$

The Figures shows a variation of the coefficient of variation at the frontier between endemic and epidemic stages, but the sense of this variation varies largely between the waves in the same country and between countries



Representation of new cases of COVID-19 (in green) and their CV coefficient of variation (in blue) during different waves of COVID-19, during the first wave in the USA (B) and during the third wave in France and Brazil (A and C).

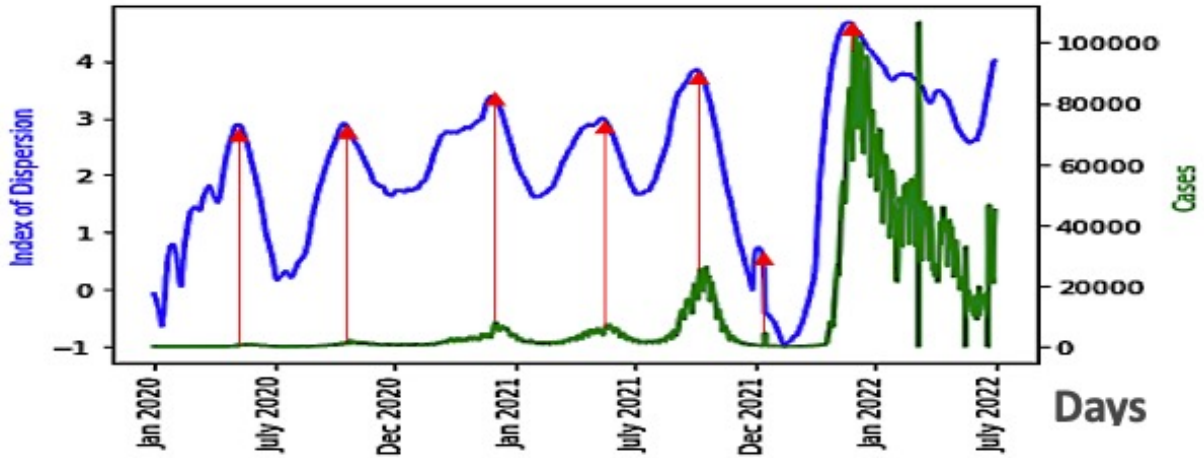
Models

The index of dispersion (*ID*) is defined by the following formula:

$$ID = \sigma^2(N_t)/E(N_t)$$

ID is equal to 0 for a constant random variable N_t and to 1 for a Poisson variable.

ID index is a good predictor of COVID-19 epidemic peaks



ID index (in blue) as predictor of the epidemic waves for Japan COVID-19 outbreak, with Daily new cases superimposed (in green).

Models

The normality index *KStest* is defined as the fitting criterion of the Kolmogorov-Smirnov test of adequation to the normal distribution having $E(N_t)$ and $\sigma(N_t)$ as respectively expectation and standard deviation of the empirical distribution of N_t.

Models

The skewness of the random variable N_t is the third standardized moment defined as

$$Skew(N_t) = E\left(\left(\frac{N_t - \mu}{\sigma}\right)^3\right).$$

Recall that the skewness verifies

$$Skew(N_t) = \frac{E(N_t^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3} = \frac{E(N_t^3)}{\sigma^3} - \left(3\frac{1}{CV} + \frac{1}{CV^3}\right).$$

The *kurtosis* is the fourth standardized moment, defined as

$$Kurt(N_t) = E\left(\left(\frac{N_t - \mu}{\sigma}\right)^4\right).$$

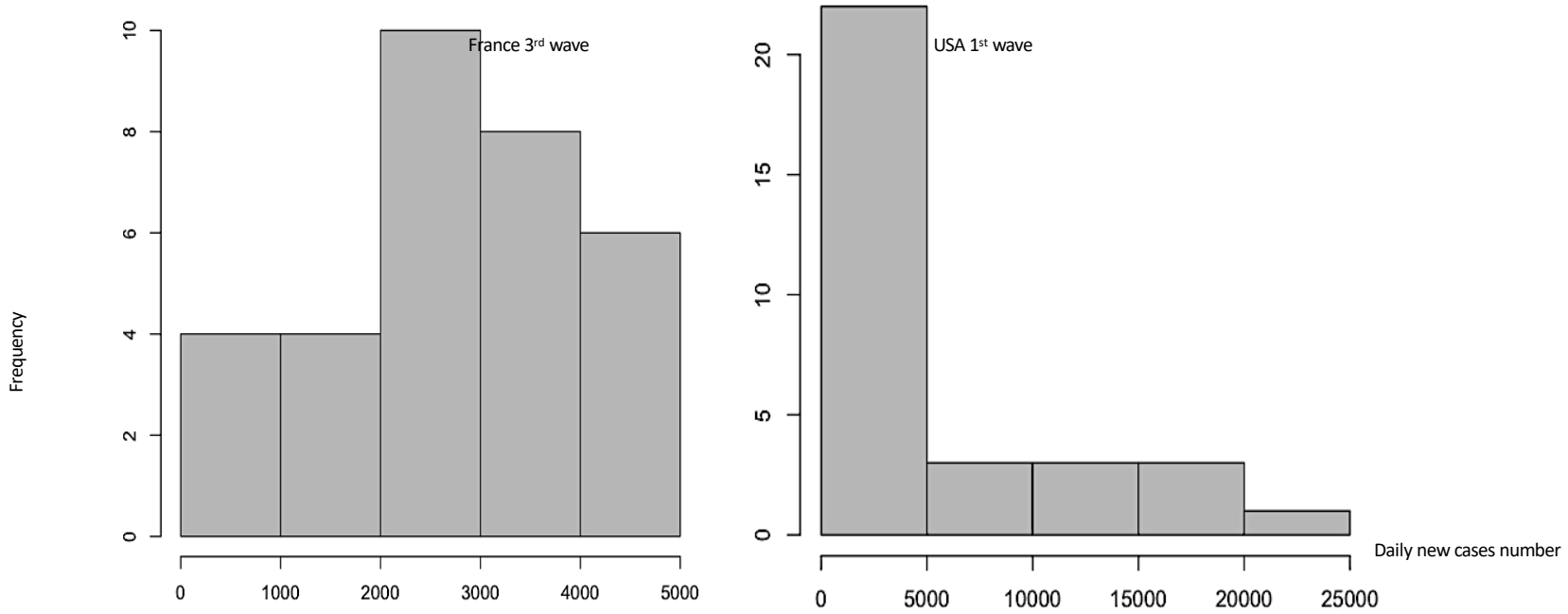
The *empirical entropy* $\mathcal{E}$ of the empirical distribution is defined as follows:

$$\mathcal{E}(N_t) = -\sum_{i=1:d \text{ with } p_i>0} p_i \log p_i,$$

where the p_i are the weights of a histogram on d value intervals of N_t.

We calculate at start of first wave in USA the entropy **E** whose value is equal to:

$$-\sum_{i=1}^{6} p_i \log p_i = p_1 \log p_1 + p_2 \log p_2 + p_3 \log p_3 + p_4 \log p_4 + p_5 \log p_5 + p_6 \log p_6 = 0.686$$



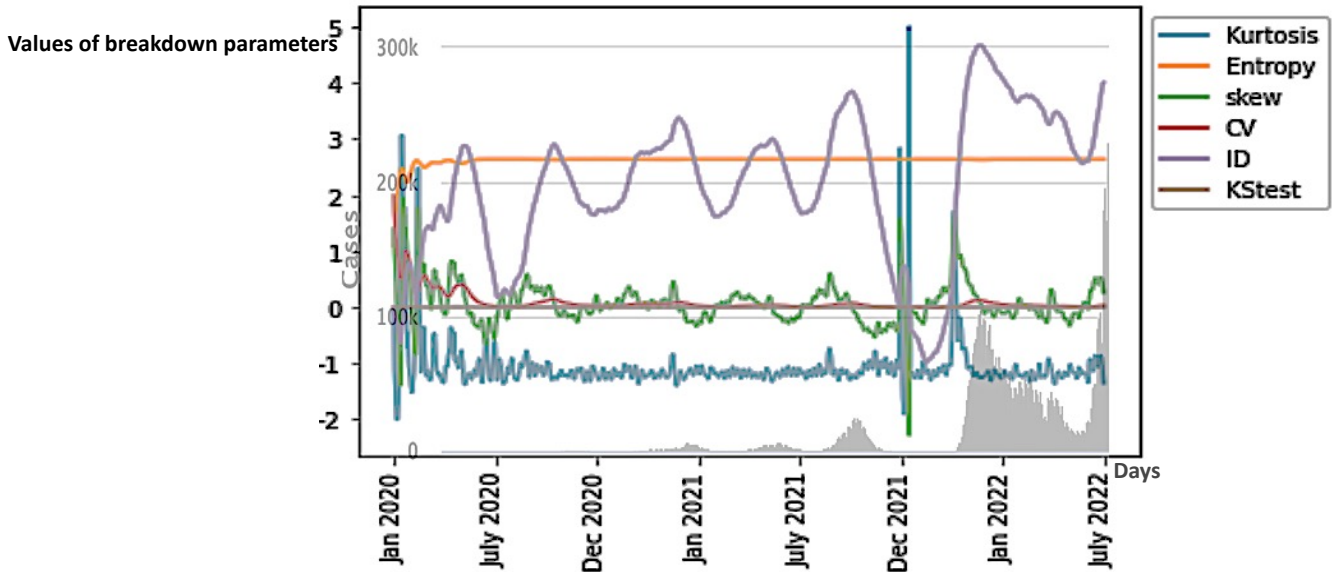
The empirical distribution of the daily new cases at the start of the USA first wave and France third wave

All indicators were calculated on same moving window respecting following rules:

- Choose the same length of moving window as for the CV calculation (14 days)
- Use the same time step as for moving the window (1 day)
- Move the window from the start to the end of the COVID-19 outbreak observed between January 2020 and July 2022.

Values of the breakdown coefficients during the first two weeks moving windows $W(i)$ ($i = 0$ to 4) for Japan during early January 2020.

i	kurtosis	entropy	skew	CV	ID	kstest	ΔID(i)
0	-0.060606	1.098612	1.392621	1.987138	-0.072551	0.000924	0.57
1	-1.100000	1.386294	0.948683	1.640825	-0.113943	0.000924	0.40
2	-1.644444	1.609438	0.596285	1.392286	-0.159701	0.000924	0.32
3	-1.916667	1.791759	0.288675	1.198289	-0.210853	0.000924	0.275
4	-2.000000	1.945910	0.000000	1.037749	-0.268845	0.000924	



Indicators and New Cases (in grey) in Japan during COVID-19 Outbreak.

Principal Component Analysis

By making a principal component analysis

we obtain the percentage of the variance explained by each principal component

$$Explain = \begin{pmatrix} 71.62 \\ 17.54 \\ 6.87 \\ 3.97 \end{pmatrix}$$

and the matrix giving the projection coefficients of the principal components

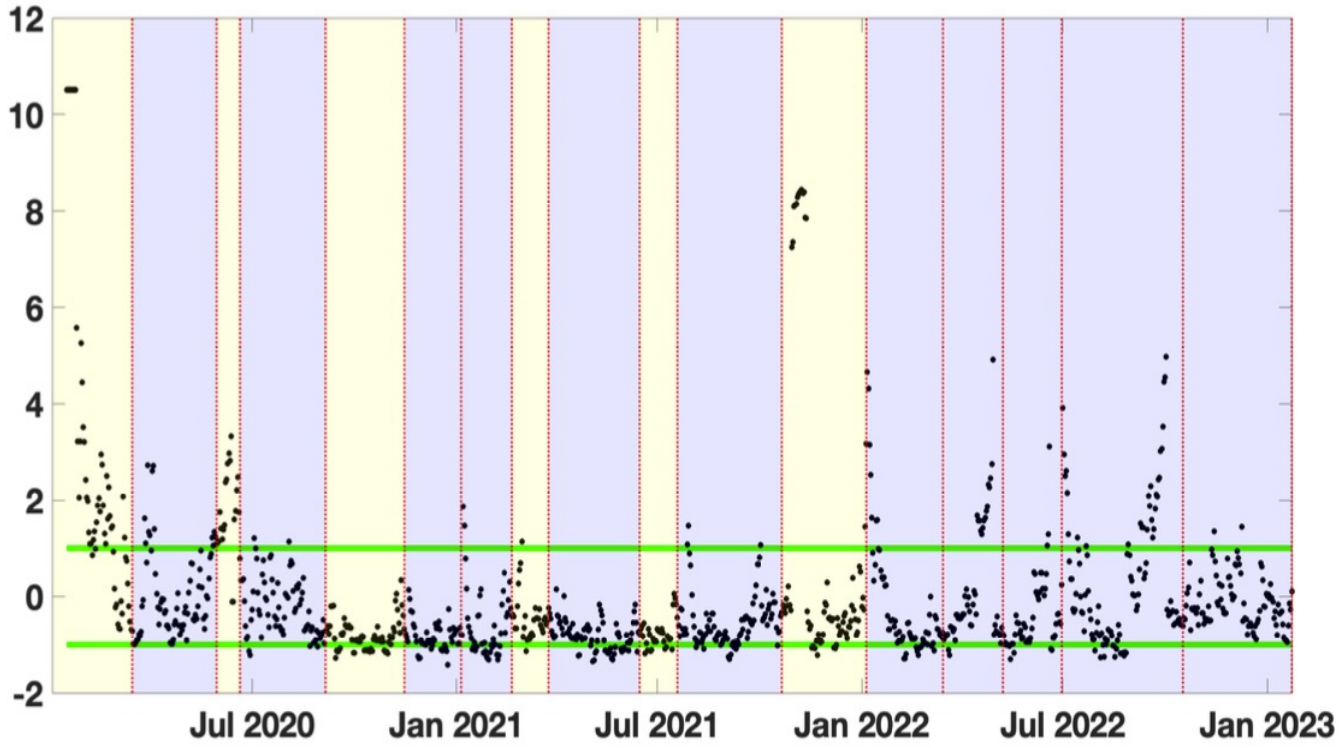
$$coeff = \begin{pmatrix} 0.5234 & -0.2243 & 0.7969 & -0.2018 \\ 0.5452 & -0.2406 & -0.2310 & 0.7691 \\ 0.5401 & -0.1760 & -0.5575 & -0.6054 \\ 0.3703 & 0.9278 & 0.0270 & 0.0358 \end{pmatrix}.$$

By using the first column of the above matrix, we deduce the first principal component

Principal Component Analysis

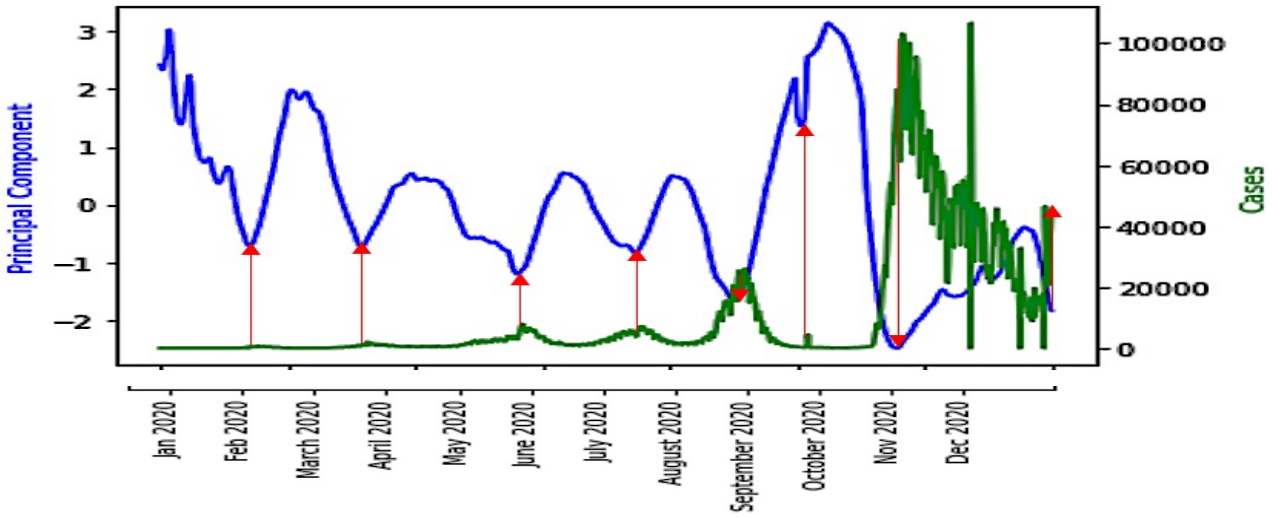
$0.52CV_s(N_t) + 0.55Skew_s(N_t) + 0.54Kurt_s(N_t) + 0.37E_s(N_t)$ which explains 71% of the variability.

We deduce that Skewness, Kurtosis, and the coefficient of variation (in decreasing order of importance) best explain the variability.



In this figure we plot the first principal component for Japan. The horizontal green lines correspond to ±1.

Principal Component Analysis

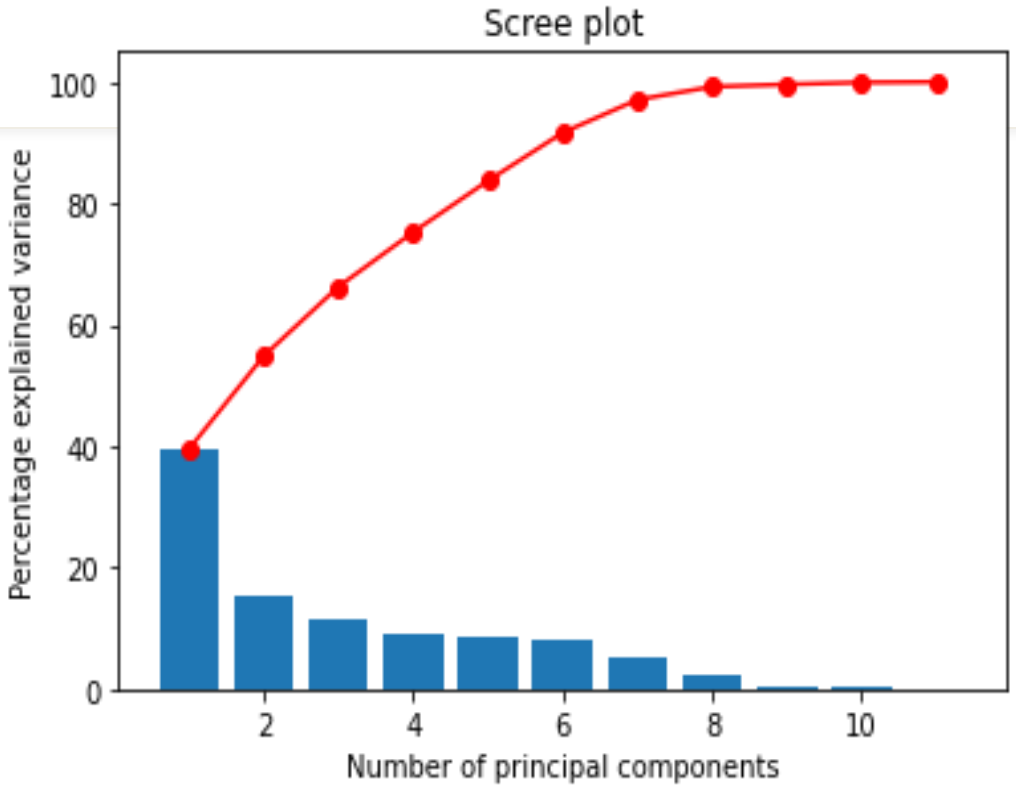


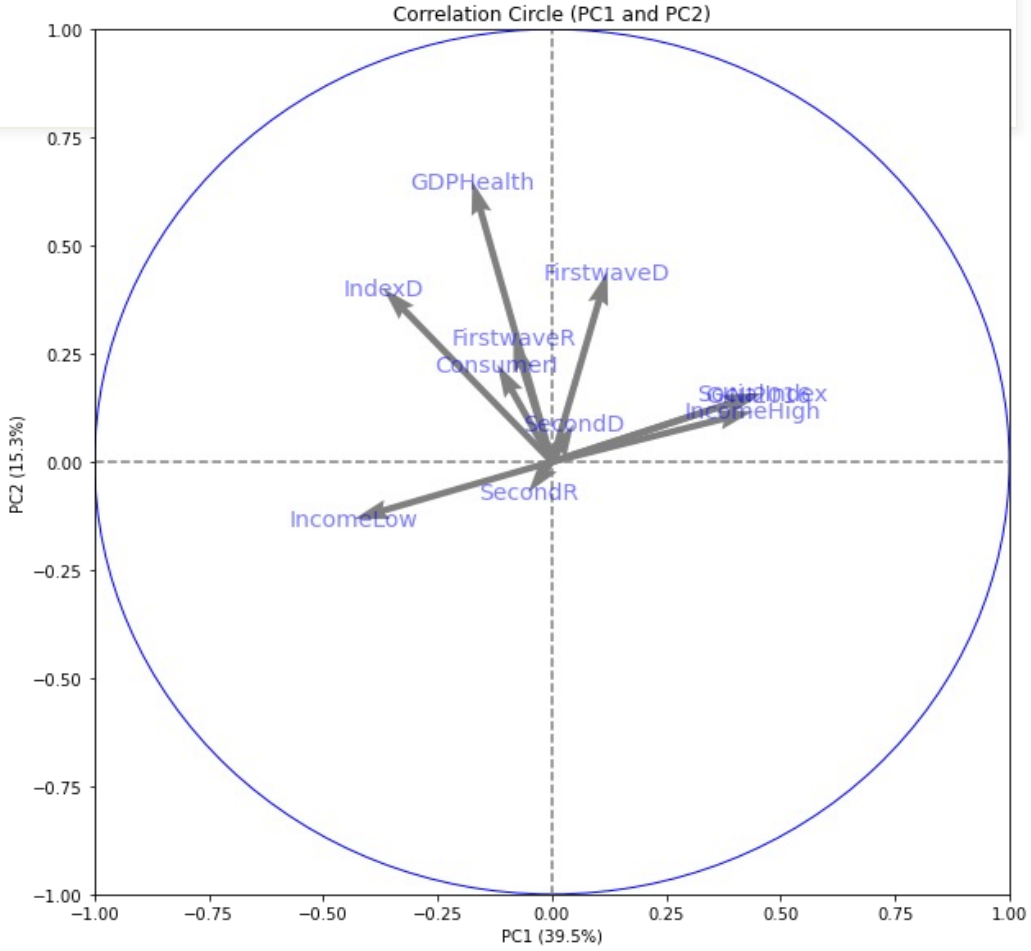
First Principal Component (blue) as predictor of COVID-19 Daily new cases waves (green) in Japan

We can conclude that among the breakdown parameters, a good predictor for epidemic waves is the first PCA component because its variations anticipate epidemic peaks.

$$PCA1 = 8.86760799\text{e-}02\ Kurt + 1.73156383\text{e-}02\ E + 1.25157924\text{e-}02\ Skew + 2.49657969\text{e-}02\ CV + 9.95518350\text{e-}01\ ID + 1.05368220\text{e-}05\ KS$$

PCA: Socio-economic-demographic influence





Gini index and social fracture index have the highest positive correlation of 0.45 and 0.46 respectively in PC 1 while percentage of GDP devoted to health expenditure and demo-economic index have the highest positive correlation in PC 2, whose values equal to 0.65 and 0.41 respectively.

Forecasting epidemic dynamics using spatial models

AIM 1

EpiMoRPH

Automated Engine

Spatial Model Construction

Synthesis with time-series data

How well do models perform?
- Model validation
- Model selection

Forecasting
(Single-model and Ensemble)

Model re-fitting

- Select a set of candidate models from the user-contributed library (AI-assist)
- Refit the models to new data
- Evaluate performance
- Produce location-specific forecasts

Use-case:
Which models are useful for my municipality, and what do those models tell me?
User uploads geo-referenced surveillance data (time-series)

AIM 2

Optimization Engine

- User-defined priorities and constraints (objective functions)
- Optimization to balance disease outcomes and equity
- Machine-assisted algorithm selection

Output:
Parameterized Model

Realistic constraints and scenarios

Output:
Sets of spatial intervention strategies ranked by likelihood of success

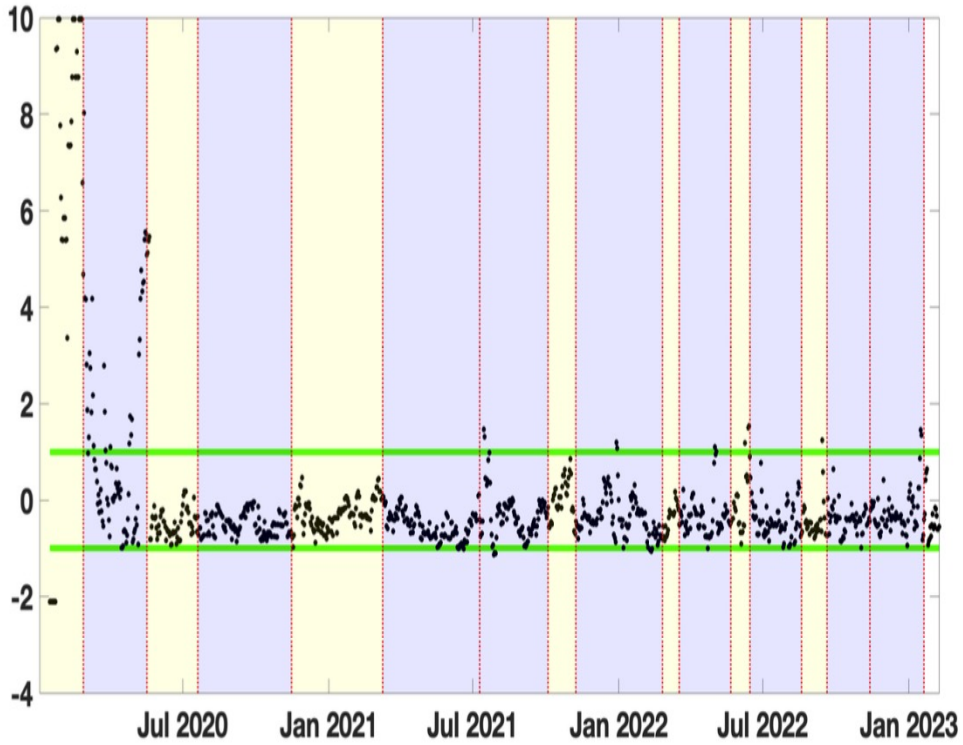
AIM 3

Evaluation & Training

Public Health Advisory Council
- Local, regional, and national stakeholders
- Quarterly meetings
- Planning, implementation, training, refinement

	Y1	Y2	Y3	Y4	Y5
AIM 1					
• Phase 1.1					
• Phase 1.2					
• Phase 1.3					
AIM 2					
• Phase 2.1					
• Phase 2.2					
• Phase 2.3					
AIM 3					
• Eval					
• Training					

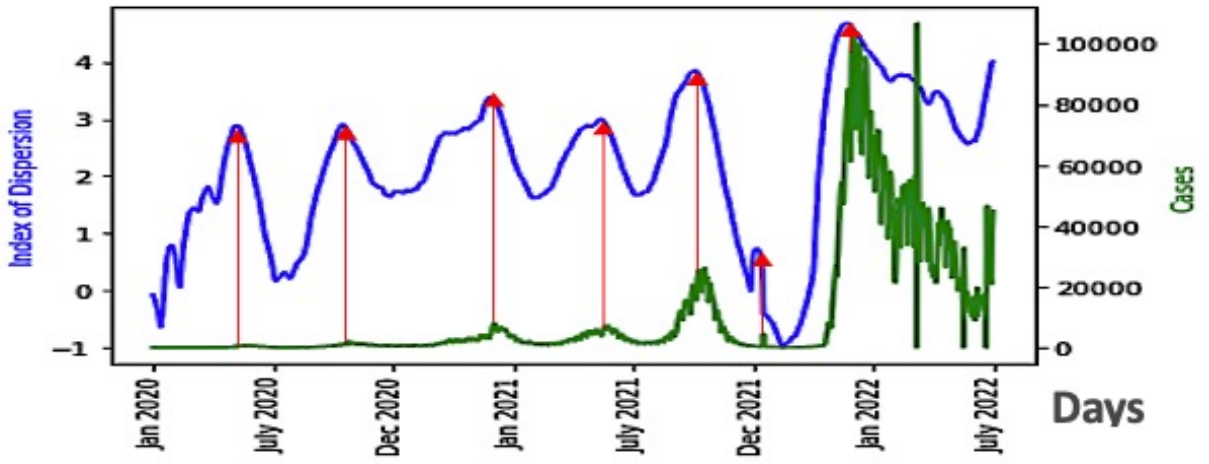
Summary and Conclusion



In this figure we plot the first principal component for France. The horizontal green lines correspond to ±1.

- We were able to build a new forecasting strategy for predicting an epidemic wave that comes after an infectious disease's endemic stationary period.
- The predictive power of the first principal component can be quantified by its performance ratio, that is, by the percentage of correct retro predictions for France obtained by fixing variation thresholds to forecast the occurrence of an epidemic wave.

Summary and Conclusion



ID index (in blue) as predictor of the epidemic waves for Japan COVID-19 outbreak, with Daily new cases superimposed (in green).

- ID index has often the main weight in the linear combination expressing PCA on the breakdown coefficients. ID waves occur in opposition of phase with PCA but predicts also well the new cases and death waves.

Scientific publications

- J. Waku, **K. Oshinubi,** U. Muhammed and J. Demongeot, Forecasting the endemic/epidemic transition in COVID-19 in some countries: influence of the vaccination, *Disease,* 2023.

- J. Demongeot, P. Magal and **K. Oshinubi**, . Forecasting the changes between endemic and epidemic phases of a contagious disease, with the example of COVID-19. *Mathematical Medicine & Biology IMA*, 2023 (Under peer review).


THANK YOU
FOR YOUR
ATTENTION

Appreciation

Many thanks to my collaborators on this project!

- Jacques Demongeot (University Grenoble Alpes, France)
- Pierre Magal (University Bordeaux, Talence, France)
- Jules Waku (University of Yaounde I, Cameroon)
- Umar Muhammed (Durham University, Durham, UK)

Any questions?