Comparative analysis of phenomenological growth models applied to epidemic outbreaks

Raimund Bürger[‡], Gerardo Chowell[§], Yissedt Lara-Díaz[‡]

[‡]Centro de Investigación en Ingeniería Matemática - CI²MA and Departamento de Ingeniería Matemática, Universidad de Concepción, Chile

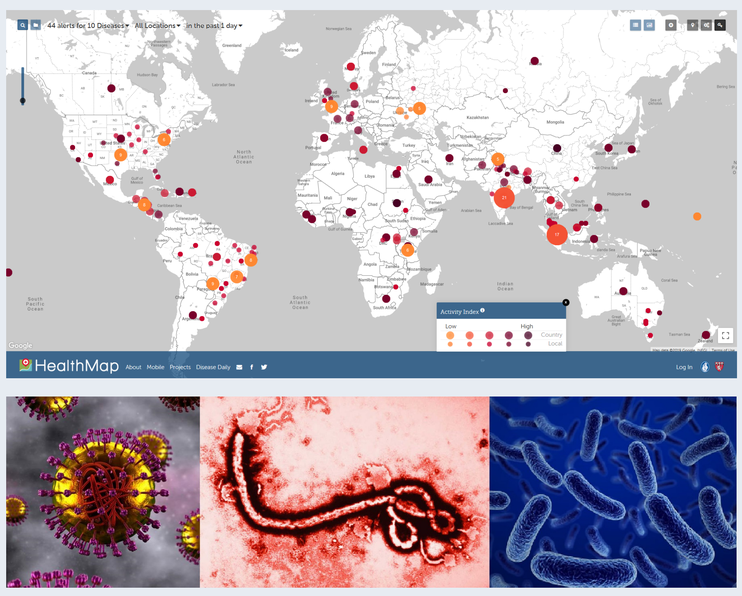
[§]School of Public Health, Georgia State University, Simon A. Levin Mathematical and Computational Modeling Sciences Center, School of Human Evolution and Social Change, Arizona State University and Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, USA

Abstract

Phenomenological models provide a mathematical framework to characterize empirical patterns without a specific basis on the physical laws or mechanism. These models often offer a simple mathematical form defined through ordinary differential equations (ODEs) that in many cases can be solved explicitly.

Also, they have been useful to characterize real epidemic trajectories. In this study, we conduct a comparative assessment using four growth models and apply them to 37 infectious disease outbreak datasets. For this goal, we estimate parameters with quantified uncertainty and assess model fit using the RMSE.

Virus and contagious disease surveillance map



Alert map (top) taked from HealthMap on Sep 19 for 10 epidemic outbreaks (Ebola, zika, dengue, measles, plague, Yellow Fever, cholera, foot and mouth, HIV/AIDS, influenza) and (bottom) virus structure pictures to measles, ebola and plague.

Mathematical models

For the following models, t denotes time, $C'(t)$ describes the incidence over time, $C(t)$ is the cumulative number of cases at time, r is the growth rate, K is the size of the epidemic, b is the exponential decay of the growth rate r and $p \in [0, 1]$ is a growth scaling parameter that indicates the kind of growth.

- Logistic Model (LM)

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = C'(t) = rC(t)\left(1 - \frac{C(t)}{K}\right)$$

- Generalized logistic Model (GLM)

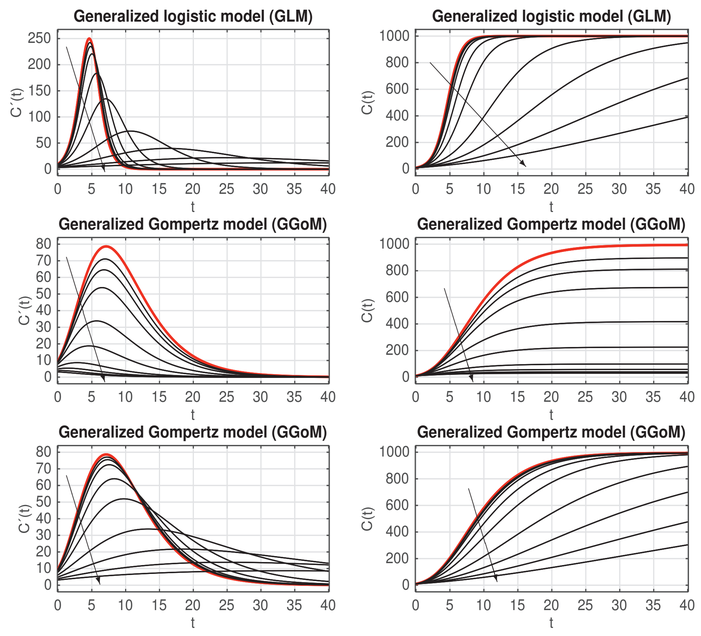
$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = C'(t) = rC^p(t)\left(1 - \frac{C(t)}{K}\right),$$

- Gompertz Model (GoM)

$$\frac{dC(t)}{dt} = rC(t)\exp(-bt)$$

- Generalized Gompertz Model (GGoM)

$$\frac{dC(t)}{dt} = rC^p(t)\exp(-bt),$$



Materials

We assess the performance of different models using multiple outbreak datasets, the RMSE error metric, parameter estimation and confidence interval generation.

- Datasets of the 37 epidemics analyzed in our study.

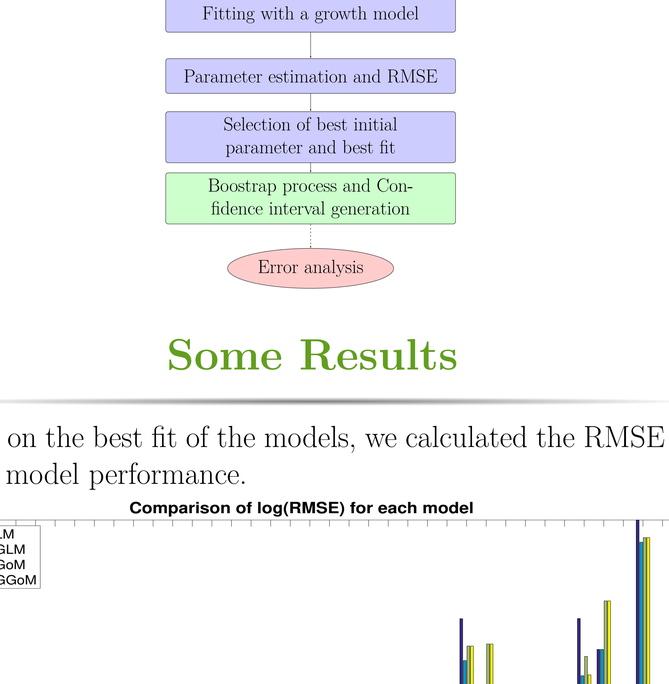
Case No.	Disease	Outbreak	Temporal resolution	Total data	Case No.	Disease	Outbreak	Temporal resolution	Total data
1	Ebola	Forecariah (GIN)	weeks	51	20	Ebola	Tonkolili (SLE)	weeks	29
2	Ebola	Gueckedou (GIN)	weeks	49	21	Ebola	Western Rural (SLE)	weeks	51
3	Ebola	Keroune (GIN)	weeks	14	22	Ebola	Western Urban (SLE))	weeks	55
4	Ebola	Kindia (GIN)	weeks	30	23	Ebola	Grand Bassa (LBR)	weeks	30
5	Ebola	Macenta (GIN)	weeks	32	24	Ebola	Congo (1976)	days	52
6	Ebola	N'Zerekore (GIN)	weeks	24	25	Ebola	Uganda (2000)	weeks	18
7	Ebola	Bomi (LBR)	weeks	33	26	Measles	London (ING) (1948)	weeks	40
8	Ebola	Bong (LBR)	weeks	17	27	Plague	Bombay (IND) (1905-06)	weeks	41
9	Ebola	G. Cape Mount (LBR)	weeks	29	28	Plague	Madagascar (2017)	weeks	50
10	Ebola	Lofa (LBR)	weeks	24	29	Smallpox	Khulna (BGD) (1972)	weeks	13
11	Ebola	Margibi (LBR)	weeks	40	30	Yellow fever	Luanda (AGO) (2016)	weeks	28
12	Ebola	Montserrado (LBR)	weeks	42	31	FMD	UK (2001)	days	121
13	Ebola	Bo (SLE) (2014)	weeks	39	32	FMD	Uruguay (2001)	days	27
14	Ebola	Kailahun (SLE)	weeks	33	33	Pandemic Flu	San Fco (USA) (1918)	days	63
15	Ebola	Kambia (SLE)	weeks	45	34	Zika	Antioquia (COL)(2016)	days	105
16	Ebola	Kenema (SLE)	weeks	39	35	VIH-AIDS	Japan (1985-2012)	years	21
17	Ebola	Kono (SLE)	weeks	30	36	VIH-AIDS	NYC (1982-2002)	years	70
18	Ebola	Moyamba (SLE)	weeks	37	37	Cholera	Aalborg (DNK) (1853)	days	105
19	Ebola	Port Loko (SLE) (2014)	weeks	54					

- The root mean square error (RMSE) is used to assess model performance and is given by,

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f(t_i, \hat{\Theta}) - y_i)^2},$$

where $\hat{\Theta}$ is the set of parameter estimates, $f(t_i, \hat{\Theta})$ denotes the best-fit model, and y_{t_i} is the time series data and n is the total number of data points.

- Parameter estimation and the confidence interval generation are obtained following Chowell et al. 2017, with the use of built-in Matlab functions LHSDESIGN and LSQCURVEFIT.

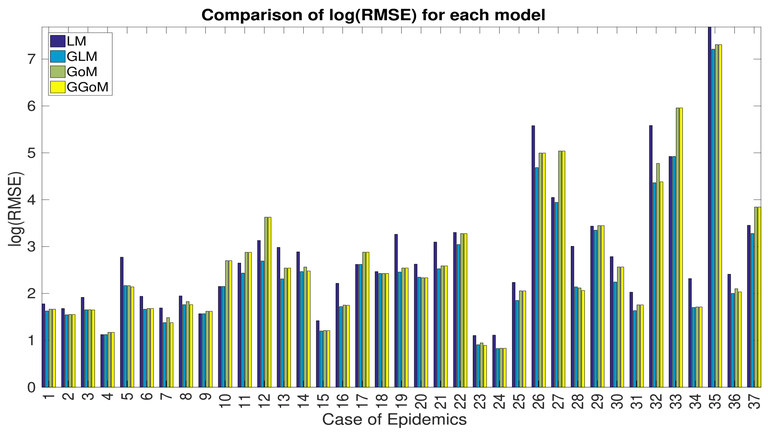


Methodology

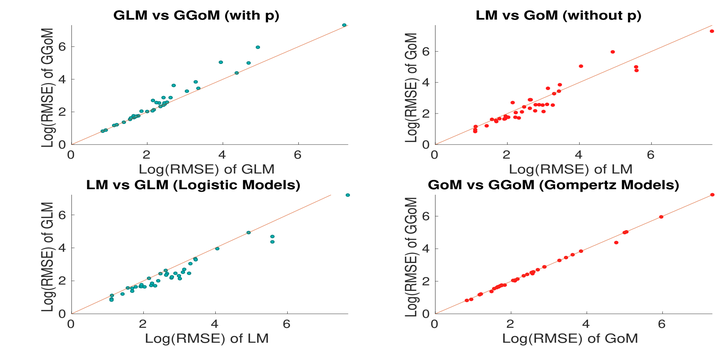
To decide which is the best model for a given outbreak and to analyze the contribution of the parameter p, the methodology consists in the following steps:

Some Results

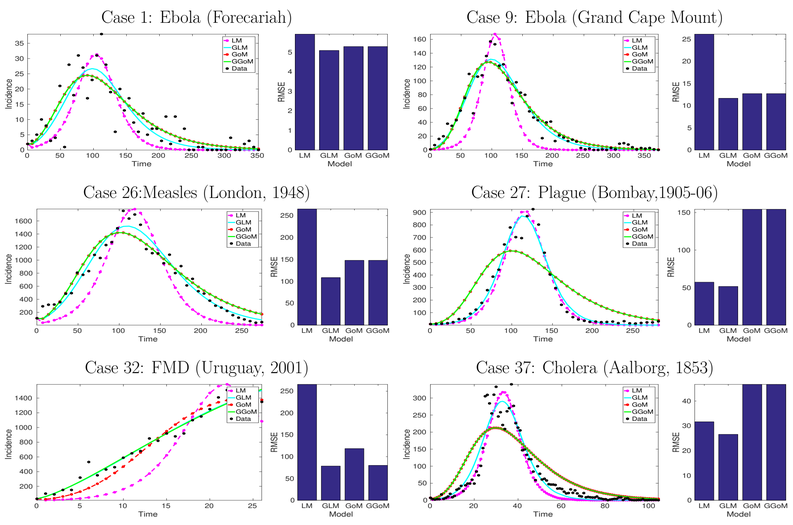
Based on the best fit of the models, we calculated the RMSE to assess model performance.



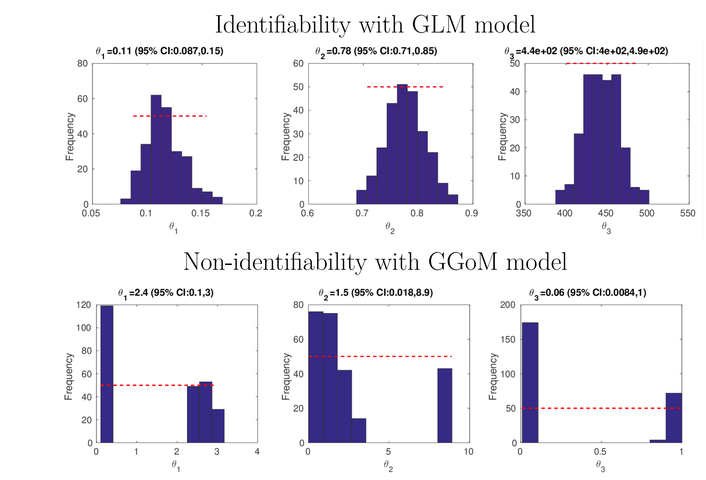
The GLM method yields the lowest RMSE in most of cases, and LM yields the larger errors.



Scatter plots Gompertz models exhibit similar dynamical behavior compared to the other models.



Figures of the models fits for representative outbreak datasets along with the corresponding RMSE values that are used to assess the quality of model fit.



Contrasting results of parameter identifiability. In the lower panel one can observe lack of parameter identifiability for the GGoM model.

Important Results

- Systematic comparison of a number epidemic outbreaks using phenomenological growth models indicates that the GLM outperfomed the other models in describing the great majority of epidemics trajectories
- Parameter p plays a much more significant role in shaping the dynamic trajectories supported by the GLM compared to the GGoM.
- Parameter estimation from both Gompertz models indicates that p is close to 1 in these models, which explains the similarity in the fits derived from these models.

References and Acknowledments

[1] G. Chowell: *Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecast*, Infectious Disease Modelling 2, (2017), 379–398.

[2] R. Bürger, C. Chowell, L. Lara-Díaz: *Comparative analysis of phenomenological growth models applied to epidemic outbreaks*, Mathematical Biosciences and Engineering,16, (2019), 4250–4273.

Support by CONICYT/PIA/AFB 170001 and CONICYT-PCHA/DoctoradoNacional/2019-21190640 is acknowledged.

Contact Information

- R. Bürger: rburger@ing-mat.udec.cl,
- G. Chowell: gchowell@gsu.edu,
- Y. Lara-Díaz: llara@ing-mat.udec.cl