Considering Viral Load in Epidemic Forecasting

Learning the Variation of Infectiousness Over Time
A Gaussian Process Approach

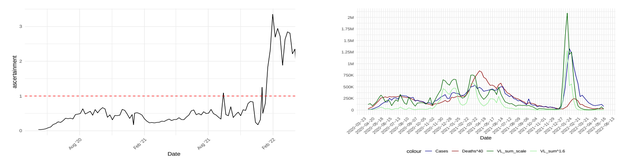
Andre Ehrlich[1*], Nikos Demiris[1], Petros Dellaportas[1,2], & Michel F. C. Haddad[3]

[1] Department of Statistics, Athens University of Economics and Business, [2] Statistical Science, University College London.
[3] School of Business & Management, Queen Mary University of London *ehrlich@aueb.gr

2nd Conference in Biostatistics & Health Analytics
July 7-9, Aegina 2025

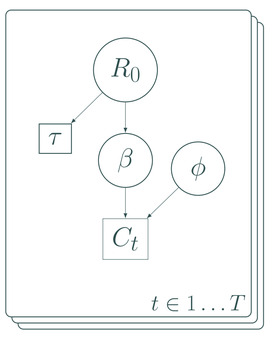
Introduction

1. Epidemic models seek to infer reproduction, prepare health system and assess efficacy of interventions.
2. Reproduction $R_t = \beta_t$ (individual infection rate) $\times$ τ (mean infectious period)
3. Reported Positive Case counts ($C_t \in 0, 1$) offer binary measure of infectiousness, symptom-driven testing bias
4. Deaths (D_t) are more trustworthy source but delayed.
5. Flaxman et al. (2020) retrospectively estimate total cases ($\tilde{C}_t$) & Ascertainment Rate ($C_t/\tilde{C}_t$)



(a) Estimated under-ascertainment (Brazil) **(b)** Various Incidence Measures

6. Viral Load $VL \in [0, 1]$ provides a continuous measure of infectiousness
7. Begin with standard epidemic model (Compartmental, SIR) using discrete-time, Chain-Binomial-like approach.



(a) Fixed Transmission Rate Model

$$
\begin{aligned}
S_t &= N_{pop} - \sum_{1}^{t} C_t \\
I_t &= \sum_{t-\tau}^{t} C_t \\
p_t &= 1 - exp(\frac{\beta I_t}{N_{pop}}) \\
1/\phi &\sim Gamma(5, 20) \\
\beta &\sim \mathcal{HN}(0, 10) \\
C_{t+1} &\sim NegBin(S_t * p_t, \phi) \\
t &= 1, \dots, T
\end{aligned}
\tag{1}
$$

Goal #1: Time-varying Transmission

Prior Work: Brownian Motion (Bouranis et al., 2022), Multiphasic Bayesian Non-parametric (Barmpounakis and Demiris, 2024).

Gaussian Process (GP) Framework

- Covariance Matrix: Specify global relationships, control smoothness, e.g. SE, Mattern

$$\mathbf{K_{SE}}(b_i, b_j|\alpha, \rho) = \alpha^2 \exp\left(\frac{-(b_i - b_j)^2}{2\rho^2}\right) \tag{2}$$

- Costly inversion $K' = \mathcal{O}(n^3)$
- Constrained, noisy observation process limits model complexity
- Stochastic Epidemic Model with log-GP prior on transmission rate, overdispersed NegBin incidence generation (**NB-GP**):

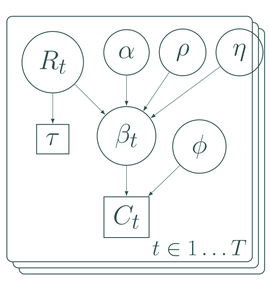
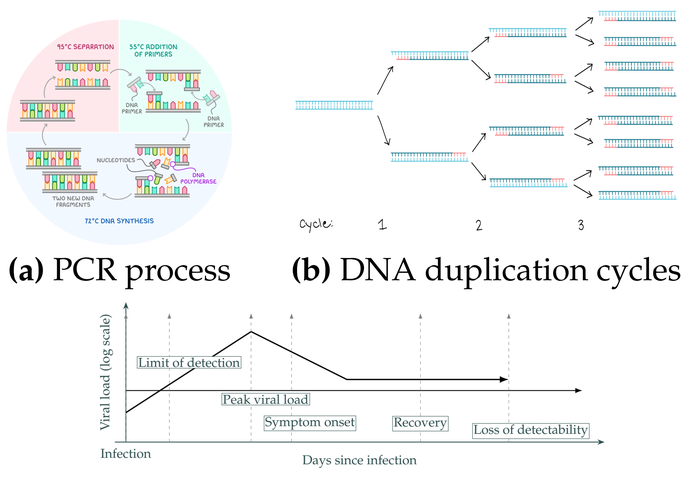


Figure 3: Infer Reproduction R_t from Cases C_t

$$
\begin{aligned}
R_t &= \beta_t \times \tau \\
C_{t+1} &\sim \text{NegBin}(S_t * p_t, \phi) \\
p_t &= 1 - \exp\left(-\boldsymbol{\beta}_t I_t / N_{pop}\right) \\
\boldsymbol{\beta} &= \exp(\boldsymbol{\lambda}) \\
\boldsymbol{\lambda} &\sim \mathcal{GP}(0, K_{SE}) \\
t &= 1, \dots, T \\
\rho, \alpha &\sim \mathcal{IG}(5, 5) \\
\phi &\sim \mathcal{G}(5, 20)
\end{aligned}
\tag{3}
$$

Goal #2: Include Viral Load



(a) PCR process **(b)** DNA duplication cycles **(c)** Viral Load Kinetics

Figure 4: How PCR measures viral load

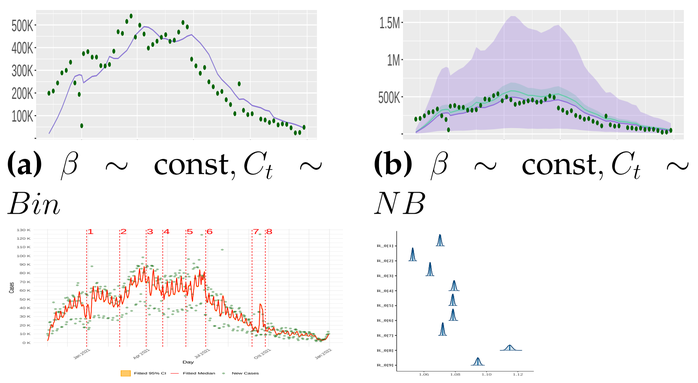
1. Hay et al. (2021) PCR Threshold value (TV) identifies growing/waning periods epidemic, small-community, consistent testing.
2. Growing epidemic has higher proportion of recent infections, ergo higher viral loads.
3. Does this hold on nationional level of testing? Inconsistent, biased testing.
4. How to scale viral load measurements? $VL = \left(1 - TV\right)^a$
5. Should we anchor this to reported positive cases? $R_t^{VL} = bR_t^C$

Case Study: COVID in Brazil

1. $N_{pop} = 212M$. By end of pandemic, 75% of Brazilians vaccinated.
2. Cumulative Measured Incidence: Reported Cases: 38.7M, Positive PCR: 250k+.
3. Total Deaths: 700k. At worst periods, 4000+ COVID-related deaths/day reported.

Infer Case Reproduction

1. The Gaussian Process model shows best calibration,
2. R_t fluctuates ≈ 1. Pandemic grows $\implies$ contacts reduce, and visa versa.
3. Spikier Cov structure (Matern) might identify interventions more precisely.



(a) $\beta \sim \text{const}, C_t \sim Bin$ **(b)** $\beta \sim \text{const}, C_t \sim NB$ **(c)** $\beta \sim$ p.w.const, $C_t \sim Bin$, Daily obs **(d)** $\beta_t \sim GP$ (Gaussian Process with NB)

Figure 5: Comparison of Fixed vs Time-Varying Infection Rate

Forecast Deaths with Viral Load & GP

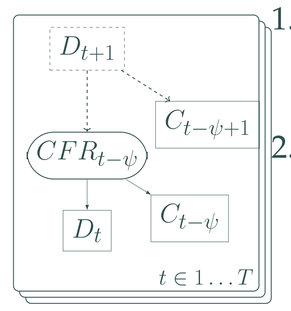


Figure 6: Death Forecast

1. Deaths useful ground truth and important epidemic quantity
2. Forecast deaths from real-time cases using observed Case-Fatality-Ratio (CFR_t),given onset-to-death time (ψ)

$$\tilde{D}_{t+\psi} = C_t \times CFR_t \tag{4}$$

3. Validate with Leave-Future-Out Prequential Analysis

Table 1: Death Forecast by Incidence Measure

	Deaths (4WA)			Incidence (1WA)		
Incidence	E	MAE	RMSE	E	MAE	RMSE
cases	4506	5947	18633	1.3e5	6.3e5	1.8e5
cases_x_VL_mean	7457	10126	29357	94376	4.6e5	1.4e5
VL_sum_scale	11270	14012	62664	4.0e5	1.9e6	5.0e5

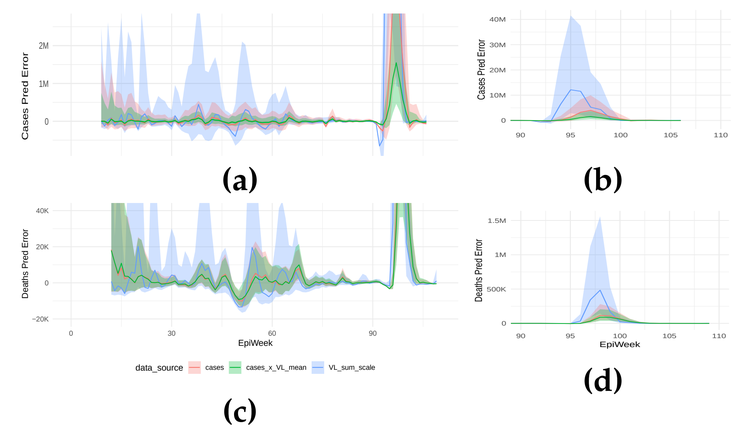


Figure 7: Forecasting Errors: 1WA Cases (top), 4WA Deaths (bottom). Zoom 3rd wave (b,d)

Key Findings

1. Inclusion of viral load shows promise to improve forecasts 1WA of incidence and 4WA of deaths,
2. GP provides natural and rich way to model time-varying dynamics within epidemic model

Further Work

1. More sophisticated scaling of VL may show more consistent improvement.
2. Make best use of other important sources of data: hospitalization, mobility.
3. Incorporate heterogeneity, social structure e.g. age stratification, spatiality
4. Analyse how social interventions may have affected reproduction