

Một mô hình gồm ba thành phần cho dịch COVID-19 và ứng dụng vào hệ thống giám sát dịch tễ

Một số chủ đề thời sự trong Toán học và ứng dụng

Khoa Toán – Cơ – Tin học, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

Viện Nghiên cứu cao cấp về Toán

Hà Nội, 30 - 31/10/2021

Hà Thành Trung
Independent researcher
TS. Toán thống kê, UF
Postdoc Y tế cộng đồng, UCF

Nguyễn Khương Tuấn
Inventor
Nhà sáng lập Onyx, STECH

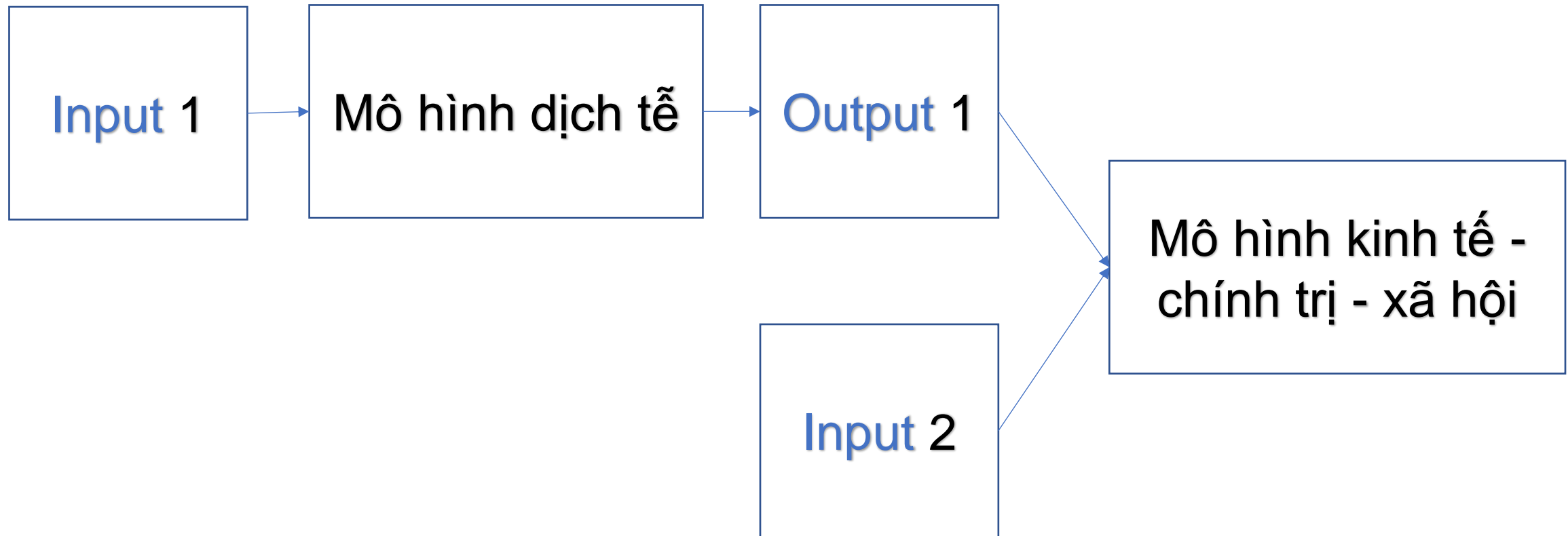


Nội dung

- Vai trò của mô hình dịch tễ
 - [Literature review](#) – thế giới
 - Việt Nam – các vấn đề thời sự
 - Ví dụ áp dụng mô hình
- Phần 1: Mô phỏng
- Phần 2: Ước lượng
- Phần 3: Dự báo
- Hệ thống giám sát dịch tễ
- Kế hoạch trong tương lai

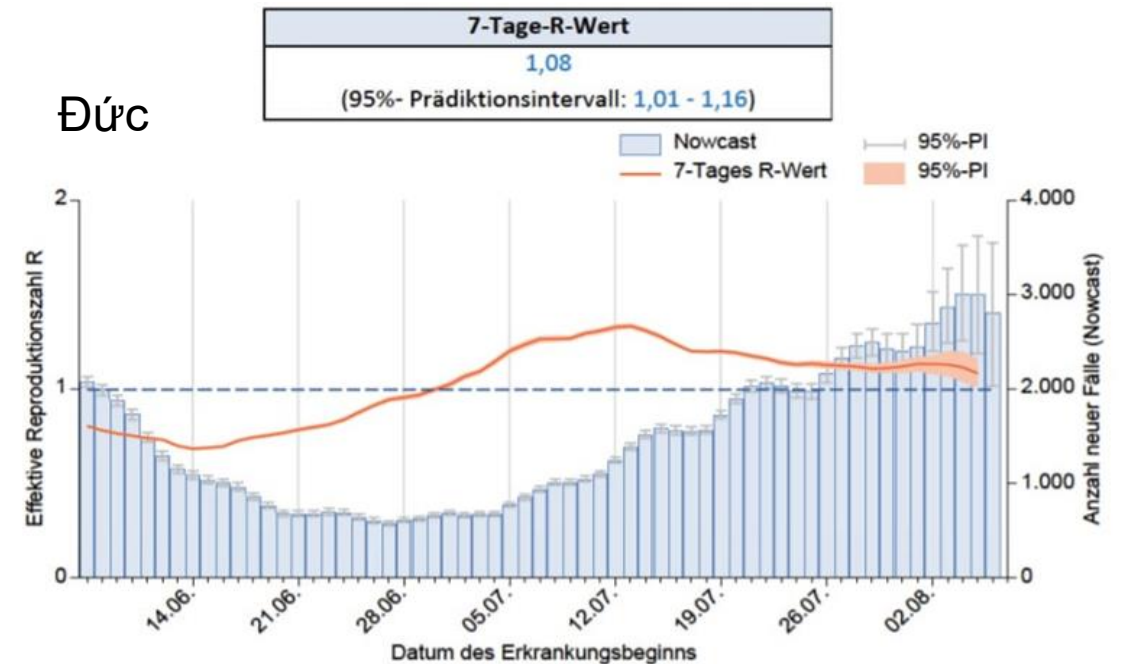
Vai trò của mô hình dịch tễ

- Gồm 3 phần: mô phỏng (phần 1), ước lượng (phần 2), và dự báo (phần 3)

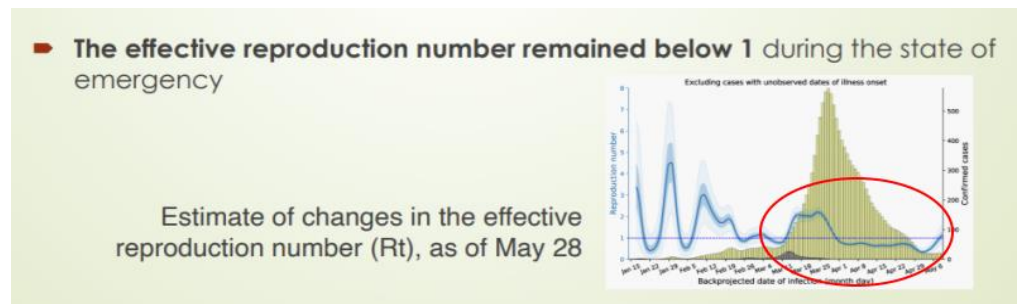


Literature review - Thế giới

- Để kiểm soát dịch, nhiều nước trên thế giới tính **hệ số lây nhiễm thực** (effective reproduction number), ví dụ như:
 - Anh ([gov.uk](https://www.gov.uk)), [Ấn Độ](#), [CovidActNow](#) (Mỹ), [Đức](#), [Israel](#), [Nhật Bản](#), [Pháp](#)
 - Hồng Kông (Trung Quốc), Na Uy, và Tây Ban Nha
- **Khoảng thời gian nối tiếp** (serial interval) là một tham số quan trọng khác liên quan mật thiết tới hệ số lây nhiễm.
- Nhiều nước sử dụng các mô hình dự báo, ví dụ:
 - [Anh](#), [Canada](#), [Đức](#), [Mỹ](#) (hơn 100 mô hình)
 - Chile, Israel, Singapore, và Tây Ban Nha
- Nhiều nước có các hệ thống giám sát dịch tễ như: Anh, Ấn Độ, Đức, Mỹ, và Úc.



Nhật Bản



Việt Nam – các vấn đề thời sự

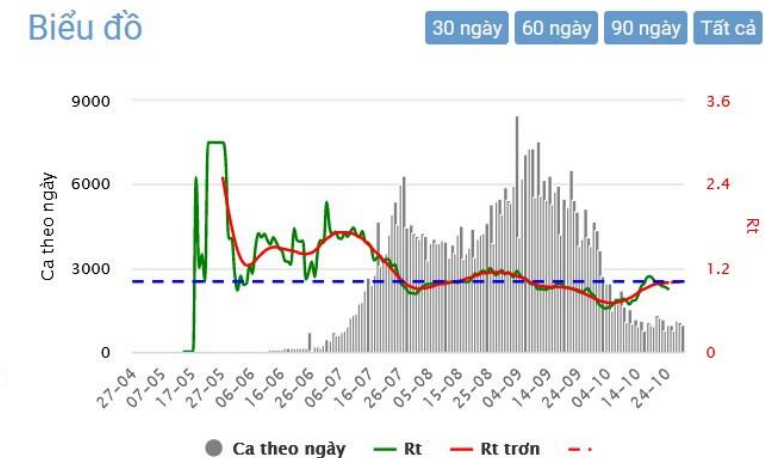
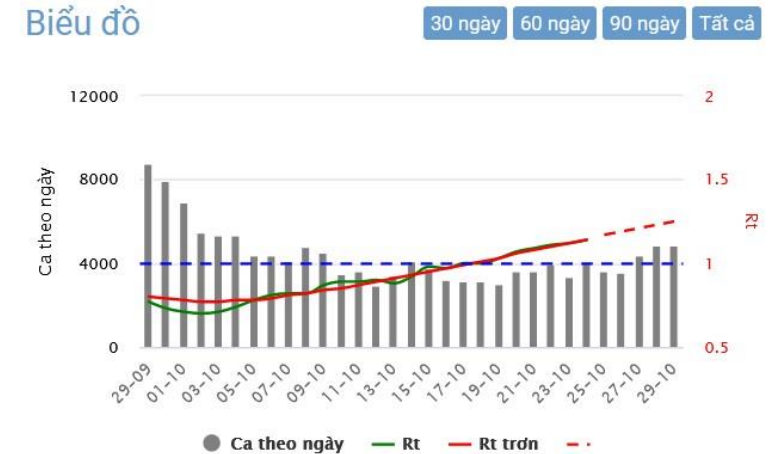
- Bộ Y tế đã đưa ra ba quyết định (QĐ) về tiêu chí kiểm soát dịch bệnh và thang đo:
 - 31/05/21: [QĐ2686](#) (chúng tôi phản biện trong [báo cáo tại Viện Toán học](#) ngày 26/08/21). Đi kèm là [bản đồ nguy cơ](#).
 - 18/08/21: [QĐ3989](#) (chúng tôi phản biện trong [bài viết](#) ngày 19/09/21)
 - 12/10/21: [QĐ4800](#) đi kèm [NQ128](#) (chúng tôi phản biện trong [bài viết](#) ngày 30/10/21).

Điểm yếu chung: Yếu định lượng nên địa phương khó áp dụng, giao địa phương tự tính gây đứt gãy liên thông, và thiếu các dẫn chứng hay nghiên cứu khoa học.

- Một số mô hình dự báo sai tình hình ở TP. Hồ Chí Minh (HCM) tháng 07/21. Nếu Chính phủ có mô hình thì TP. HCM có lẽ không phải đặt hàng gấp gáp.
- Các lãnh đạo cấp cao tuyên bố những thông tin sai lệch như “[xóa chu kỳ lây nhiễm](#)” hay “chu kỳ ... chủng mới chỉ từ [4 - 8 tiếng](#)”. (Đáp án: 4 ngày).
- Dường như Chính phủ Việt Nam không có mô hình nào nên có đơn đặt hàng gần đây [trên báo chí](#).

Ví dụ áp dụng hệ số lây nhiễm

- Ví dụ 1 (Ảo giác số ca đi xuống): Số ca có xu hướng đi xuống 3 tuần thì tình hình đang tốt lên?
- Ví dụ 2 (Ảo giác số ca nhập viện/tử vong đi xuống): Số ca tử vong và/hoặc nhập viện đi xuống 3 tuần thì tình hình đang tốt lên?
- Ví dụ 3 (Ảo giác kiểm soát dịch): Chúng tôi đã chỉ ra vì sao TP. HCM chưa bao giờ thực sự kiểm soát được dịch trong [bài viết](#) ngày 09/07/21. Tình hình cứ xấu dần đi một cách chậm rãi rồi bùng nổ (có thể do một số vụ siêu lây nhiễm) theo cấp số nhân và nhanh chóng vượt ngoài tầm kiểm soát.
- [Tháng 04/20](#), chúng tôi đã cảnh báo cần đo đạc hệ số lây nhiễm cho mọi vùng.
- Đánh giá tác dụng làm giảm $R_{e,t}$ của các biện pháp quần chúng (mang tính rộng khắp như đóng cửa trường học) và các biện pháp nhà nước (mang tính hữu hạn về nhân/nguồn lực như cách ly) như thế nào? Nếu các biện pháp nhà nước có khả năng quá tải, cần đảm bảo các biện pháp quần chúng là đủ để giảm $R_{e,t} < 1$.



Phần 1 – Mô phỏng giai đoạn đầu

- Vấn đề: khi chưa có dữ liệu thì ta có thể làm gì?
- Giải pháp: Lịch sử. Xem tất cả các nước.
- Điều kiện cần: Khả thi khi một chủng COVID-19 đã xảy ra ở nước khác. **Hệ số lây nhiễm cơ bản** R_0 và **khoảng thời gian nối tiếp trung bình** T đã được ước lượng.
- Ý nghĩa: **1 ca bệnh tạo ra R_0 ca bệnh trong khoảng thời gian T** (tương tự GDP được tính trong 1 năm).
- Trong giai đoạn đầu nT ngày, khi $R_0 > 1$, 1 cá thể bệnh sẽ sinh ra tổng số cá thể bệnh (trung bình):

$$\frac{R_0^{n+1} - 1}{R_0 - 1}$$

- R_0 cho vùng thưa dân cư sẽ nhỏ hơn R_0 cho thành phố.
- R_0 là không đủ nếu thiếu T . Cặp ($R_0 = 5, T = 4$) sẽ khiến số ca tăng nhanh hơn cặp ($R_0 = 25, T = 8$).

$$T = 4$$

Chủng COVID-19	R_0	Tổng số ca sau 20 ngày
SARS-CoV-2	2	63
Alpha	3.4	643
Delta	5	3906

Chính phủ nhiều nước như Anh có [báo cáo mỗi 2 tuần](#) về các chủng

Áp dụng

Áp dụng 1: ước lượng kích thước ổ khi phát hiện ổ dịch.

- Input: Ngoài R_0 và T , ta cần thêm 2 tham số: xác suất phát hiện ca bệnh p và thời gian trung bình từ khi mắc bệnh đến khi được phát hiện τ .
- Thuật toán:
 - Bước 1: Với xác suất 95%, tính số ca m tối thiểu để phát hiện ra ổ.
 $(1 - p)^m < 5\%$
 - Bước 2: Quy đổi m ra số ngày.
- Ví dụ với $p = 10\%$, $\tau = 7$:

Chủng COVID-19	R_0	Số ngày để có ít nhất 29 ca	Số ngày cho tới khi đi viện	Tổng số ca khi phát hiện ổ
SARS-CoV-2	2	16	23	63 đến 127
Alpha	3.4	12	19	188 đến 643
Delta	5	8	15	156 đến 781

- Ví dụ: Áp dụng ở Đà Nẵng vào 07/2020.

Áp dụng 2: ra chính sách truyền thông kêu gọi người có dấu hiệu dịch tễ đi xét nghiệm sớm.

Áp dụng 3: dự báo ngăn ngày khi mới có ổ dịch.

Phần 2 – Ước lượng

Đặt vấn đề

- Chúng ta chỉ có:
 - Input: số ca được phát hiện mỗi ngày (**reported cases**)
 - Có các ca xâm nhập (**imported cases**)
- Cần giải quyết:
 - Có các biện pháp can thiệp như giãn cách, khẩu trang, cách ly .v.v tại các thời điểm khác nhau
- Ta xem xét một phương pháp cân bằng giữa sự **hợp lý** và tính **giải được**:
 - Xuất phát từ các khái niệm cơ bản để lựa chọn mô hình **time-since-infection** (sử dụng mô hình SIR để mô tả nó).
 - Bám sát câu hỏi để tìm loại **reproduction number** phù hợp
 - Những vấn đề khi áp dụng

Nhìn nhận vấn đề

Vấn đề có sự tham gia của 3 yếu tố cơ bản:

- Môi trường,
 - Đối tượng,
 - Quá trình tăng/giảm số lượng của đối tượng.
- Môi trường là nơi đối tượng có thể tồn tại (người, dơi, v.v.). Nó thường **hữu hạn**. Ở đây giới hạn trong loài người.
 - Đối tượng có thể gồm nhiều loại loài/hạt/v.v. cùng một lúc như trong hệ sinh thái. Ở đây là vi-rút.
 - Trường hợp của ta, quá trình tăng/giảm vi-rút được đơn giản hóa thành quá trình tăng/giảm số cá thể bệnh. Tuy nhiên ta vẫn mô hình hóa nồng độ vi-rút trong cá thể bệnh (**viral load**).
 - Quá trình tăng giảm số ca bệnh được tính qua các hệ số lây nhiễm.
 - Các ví dụ: quả bom, xã hội và trái đất, vi-rút và xã hội.
 - **Giám đốc** CDC Mỹ: “con người đang chiến đấu với nhau chứ không phải chiến đấu với kẻ thù chung là vi-rút”

Effective reproduction number

- **Effective reproduction number** $R_e(t)$: the average number of secondary infections caused by a typical primary infection at time t [1]
- Đặc điểm:
 - Liên hệ giữa hai **thế hệ** (primary, secondary). Hệ quả là tập cho **primary** và tập cho **secondary tách rời**.
 - Chỉ lây **một** lần duy nhất (hai thế hệ liên tiếp)
 - Chỉ dễ hiểu khi các cá thể có cùng R_e (**đồng nhất**). Khi đó để tính R_e , ta chỉ cần số ca trong mỗi thế hệ.
- Một số vấn đề:
 - Có nhiều thế hệ tại cùng một thời điểm.
 - Nếu không có quy luật cho $R_e(t)$, thì ta có n biến trong n ngày. Dữ liệu đầu vào chỉ có n số cho n ngày. Do đó $R_e(t)$ phải là biến duy nhất thì bài toán mới **giải được**. Ta xét một mô hình thỏa mãn điều này.

Ba giả thiết cơ bản

- (Rời rạc) Do dữ liệu là số ca mới mắc mỗi ngày, giả thiết thời gian là ngày ($t \in \mathbb{N}$). Điều này giúp đơn giản hóa mô hình.
- (Độc lập) Để giải quyết vấn đề có nhiều thể hệ tại một thời điểm, giả thiết quá trình truyền bệnh của mỗi cá thể là độc lập. Như vậy quá trình truyền bệnh của một cá thể phụ thuộc thời điểm nó mới mắc bệnh và quá trình truyền bệnh của nó, chứ không phụ thuộc vào thể hệ của nó. Từ đó ta có thể gán thể hệ cho các ca bệnh và tìm các đại lượng tương tự R_e để đánh giá dịch.
- (Đồng nhất) Để vấn đề có thể giải được với dữ liệu chỉ là số ca mới mắc mỗi ngày, giả thiết quá trình truyền bệnh là như nhau cho các cá thể mới mắc bệnh tại cùng một thời điểm. Như vậy sự loại bỏ khỏi quá trình truyền bệnh cũng đồng nhất. Nếu không có tính đồng nhất, thì cần thêm thông tin về từng nhóm và sự truyền bệnh giữa các nhóm.
- Ta sẽ tìm cách dựng lại quá trình bệnh lan truyền từ dữ liệu (ở mức đủ để trả lời câu hỏi, chứ không biết ai lây cho ai).

SIR (thời gian rời rạc)

Nhắc lại mô hình SIR [2]:

- **Giả thiết 1:** thời gian là ngày ($t \in \mathbb{N}$) và sự kiện gặp gỡ cũng như lây bệnh chỉ xảy ra tại các thời điểm cách nhau đúng 1 ngày
- Tập vật chủ được chia thành 3 tập thay đổi theo thời gian:
 - \mathcal{S}_t : (**susceptible**) lành không có miễn dịch, có thể nhiễm bệnh
 - \mathcal{I}_t : (**infective**) bệnh, có thể lây bệnh cho cá thể trong \mathcal{S}_t
 - \mathcal{R}_t : (**removed**) không liên quan đến quá trình truyền bệnh (ví dụ như cách ly (**isolation**), có miễn dịch, khỏi bệnh/chết)
- S_t, I_t : lần lượt là số cá thể trong \mathcal{S}_t và \mathcal{I}_t
- **Giả thiết 2:** Tổng số cá thể $N > 0$ không đổi

- **Giả thiết 3:** Một cá thể gặp $m > 0$ cá thể bất kì 1 ngày. Các sự kiện gặp gỡ là độc lập.

Để lây bệnh cần cá thể trong \mathcal{J}_t gặp cá thể trong \mathcal{S}_t ngày $t + 1$.
 Một cá thể trong \mathcal{J}_t gặp $m \frac{S_t}{N}$ cá thể trong \mathcal{S}_t trong ngày.

- **Giả thiết 4a:** Tỷ lệ lây bệnh là $0 < p < 1$ cho mỗi gặp gỡ giữa \mathcal{J}_t và \mathcal{S}_t .

Gọi \mathcal{V}_t là tập cá thể mới mắc bệnh tại t . V_t là số cá thể trong \mathcal{V}_t .

$V_{t+1} = (\text{số cá thể } \mathcal{J}_t) * (\text{số cá thể } \mathcal{S}_t \text{ mà 1 cá thể } \mathcal{J}_t \text{ gặp}) * (\text{tỷ lệ lây})$

$$= I_t \left(m \frac{S_t}{N} \right) p = \frac{mp}{N} S_t I_t =: \beta S_t I_t, \text{ với } \beta = \frac{mp}{N}.$$

Ta có

$$V_{t+1} = \beta S_t I_t$$

- βS_t là tốc độ truyền bệnh tại thời điểm t
- Chú ý: ta xét m và p thay đổi theo thời gian sau này. Ta không xét phương trình thứ 2 của SIR vì ta không dùng nó ở đây.

$$V_t = \beta S_t I_t = \frac{mp}{N} S_t I_t$$

Các cách làm giảm số ca mắc mới (phù hợp kinh nghiệm thực tế):

- Giảm tỷ lệ lây bệnh p : khẩu trang, rửa tay, .v.v.
- Giảm số gặp gỡ m : giãn cách xã hội, quãng đường di chuyển, .v.v.
- Giảm số cá thể không miễn dịch S_t : vắc-xin
- Giảm số ca bệnh I_t : cách ly, nhập viện
- Tăng N ?

Mô hình **time-since-infection**

Cả SIR và **time-since-infection** với tham số không đổi theo thời gian là các trường hợp riêng của Kermack-McKendrick (KM) model. Các giả thiết 1-3 vẫn như SIR [2, 3]. **Time-since-infection** cho phép nhiều tốc độ lây thay vì chỉ một trong SIR.

- **Giả thiết 4b**: Khi 1 cá thể **mới** mắc bệnh tại $t \in \mathbb{N}$, nó chỉ lây cho cá thể khác trong T ngày tiếp theo. Tỷ lệ lây bệnh là p_k cho mỗi gặp gỡ sau k ngày (gặp gỡ tại thời điểm $t + k$ với $k = 1, \dots, T$), trong đó $\sum p_k > 0$. **Để tiện cho trình bày, chọn $T = 4$.**
- **Giả thiết 5**: Không cá thể bệnh nào bị loại bỏ trong vòng T ngày kể từ khi mới mắc bệnh.
- $S_t, I_t,$ và R_t được biểu diễn lại trên trục thời gian. Chú ý $I_t = V_t + \dots + V_{t-3}$.

Xác định đại lượng

- **Instantaneous reproduction number** R : số ca $T = 4$ ngày trước lây cho bao nhiêu ca hôm nay với giả thiết số ca mỗi ngày trong 4 ngày đó như nhau.
- **Case reproduction number** R_c : 1 ca hôm nay lây cho bao nhiêu ca trong tương lai.

$$R_c(t) := R_{t+1}w_1 + \dots + R_{t+4}w_4.$$

- Hệ số lây nhiễm R_T ($I_t > 0$)

$$R_T(t) = \frac{I_{t+T}}{I_t}.$$

Chúng tôi chọn R_T (dịch chuyển 4 ngày so với [4]) vì nó không cần w , dễ hình dung, và dễ suy luận.

- Do chu kỳ hành vi 7 ngày của xã hội, chính phủ Đức thay đổi công thức một chút và chúng tôi thay đổi theo.

Ngẫu nhiên hóa

- Khi dữ liệu đầu vào V_t không chính xác (**missing data**, độ trễ khi thu thập dữ liệu, .v.v.), ta có thể giả thiết một phân phối cho toàn bộ dữ liệu ($V_0, V_1, \dots, V_t, \dots$).
- Có rất nhiều cách khác nhau để ngẫu nhiên hóa. Ví dụ như tùy vào việc thu thập dữ liệu [4], hoặc sử dụng một phân phối phổ biến như quá trình **Poisson** cho dữ liệu đếm [5].
- Nếu giả thiết không phù hợp/**robust**, thì **confidence/credible interval** chỉ tạo ảo giác và có thể phản tác dụng.
- Do quan hệ giữa V và R_T là 1-1, chúng tôi coi như R_T là ngẫu nhiên. Tìm hiểu các nước có dữ liệu tốt như Đức, chạy mô phỏng đơn giản cho 1 số tình huống (như T hơi lệch, chu kỳ hành vi xã hội, thay đổi chiến thuật xét nghiệm), chúng tôi lựa chọn hồi qui phi tham số để làm trơn R_T .

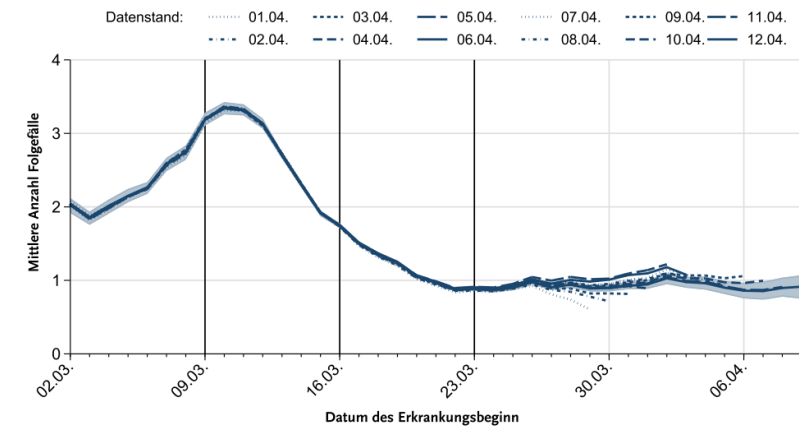


Abb. 5 | Vergleich der Schätzung der effektiven Reproduktionszahl R für eine angenommene Generationszeit von 4 Tagen zu unterschiedlichen Datenständen. Die schwarz hervorgehobenen vertikalen Linien kennzeichnen den Start der in Tab. 1 (S. 15) genannten Maßnahmen am 9. März, 16. März und 23. März 2020.

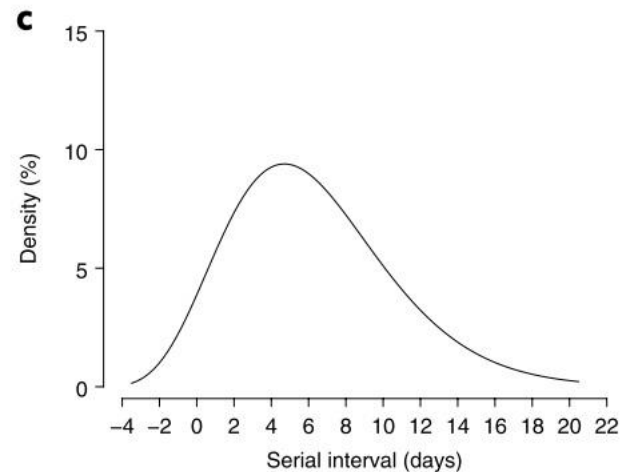
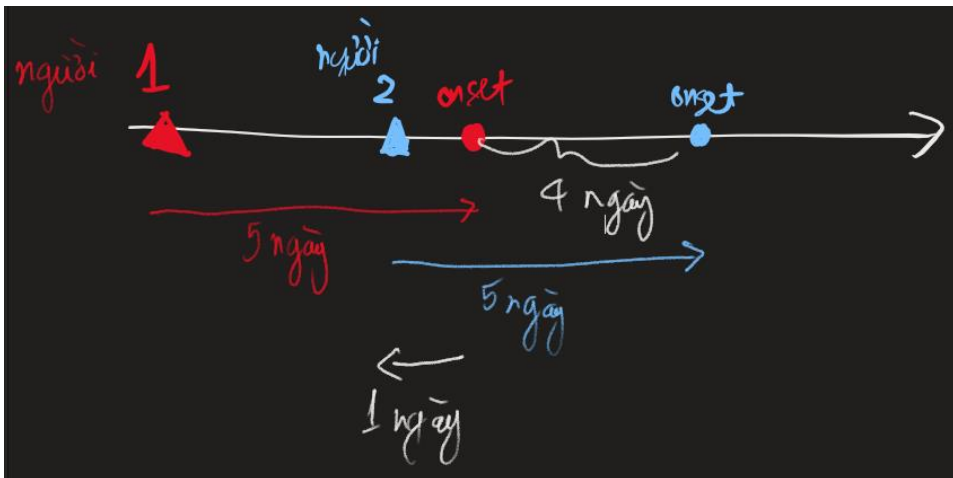
Ứng dụng

- Với một cá thể ta có các thời điểm

Infection of a virus < onset of symptoms < removal

Incubation time: khoảng thời gian giữa **infection of a virus** và **onset**

- Khi hai cá thể lây cho nhau, ta có các khoảng thời gian
 - **generation time:** giữa 2 **infections**
 - **serial interval (SI):** giữa 2 **onsets**, có thể âm (hình dưới) [6]
- Thực tế ta chỉ quan sát được **onset of symptoms**. Nên V_t sẽ đếm **onset**, và w sẽ cho **SI**.
- Vì các hệ số lây nhiễm được tính cho **onset** và SI, ta cần chuyển đổi nó sang hệ số lây nhiễm **thật** thông qua các giả thiết khác (như **incubation time** độc lập SI, .v.v. [1]).

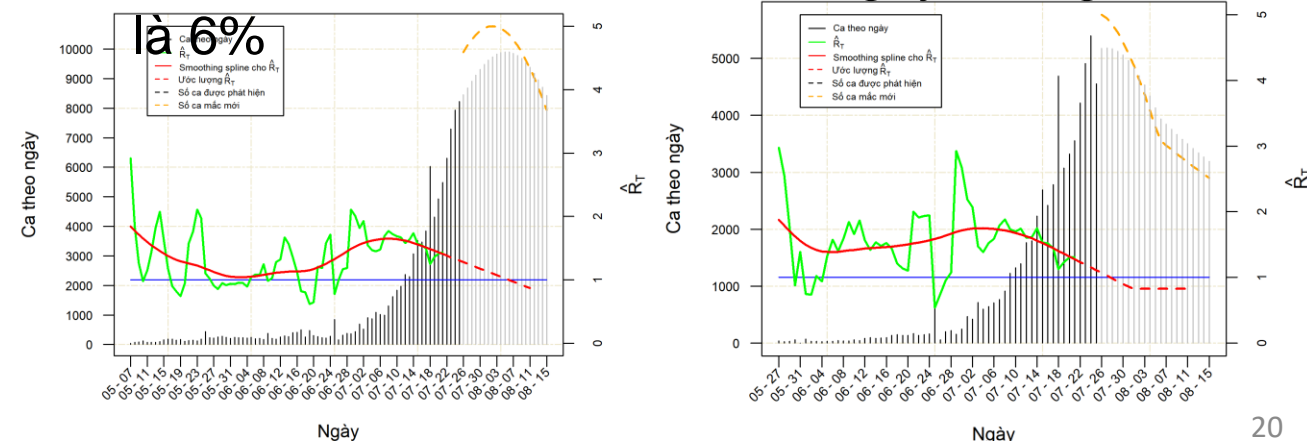


Phần 3 – Compartmental model

- Chúng tôi chọn SEIR. Kí hiệu t_0 là thời điểm xuất phát.
- Nếu sử dụng **compartmental model** khác với nhiều **compartment** hơn thì cần ước lượng nhiều tham số hơn. Các thủ tục tương tự. Công thức cho R_e được tính qua **next-generation method** [7].

Tham số	Giá trị	Lý do
N : Tổng dân số VN	98 triệu	Xấp xỉ
E_0 : Số ca bệnh chưa được phát hiện hôm t_0	Tổng số ca phát hiện từ $t_0 + 1$ đến $t_0 + 4$	Đây là khoảng thời gian đa số những ca này được phát hiện
I_0 : Số ca bệnh được phát hiện trong vòng T ngày cho tới t_0	Tổng số ca phát hiện từ $t_0 - 3$ đến t_0	
$Remove_0$: Số ca không liên quan quá trình lây bệnh	0	Không ảnh hưởng quá nhiều tới mô hình
γ : Thời gian ủ bệnh trung bình	4 ngày	T
δ	4 ngày	T
R_e	Hồi quy qua $R_T(t_0)$	
β	Tính từ các tham số trên	

- Chạy tự động cho VN/tỉnh/huyện/xã
- Chúng tôi không tính khoảng tin cậy vì khoảng tin cậy cho spline ngoài khoảng dữ liệu rất lớn
- Chúng tôi phát triển một thuật toán dự báo khác bằng cách tính tốc độ tăng số ca theo ngày
- Sai số của một số dự báo 21 ngày trước là 6%



Hệ thống giám sát

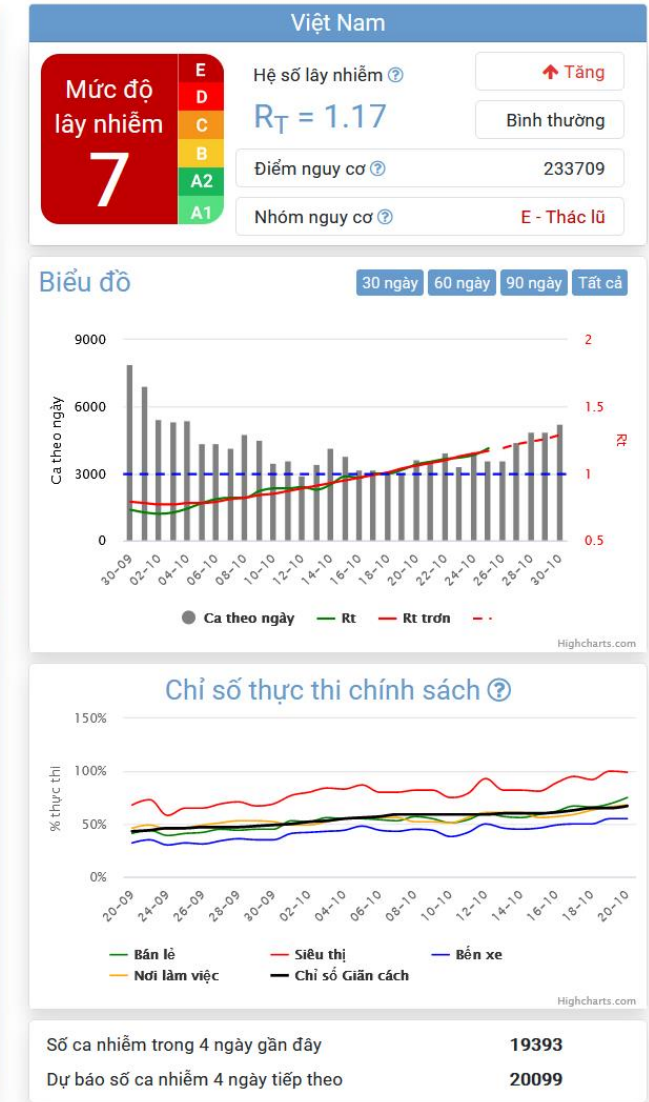
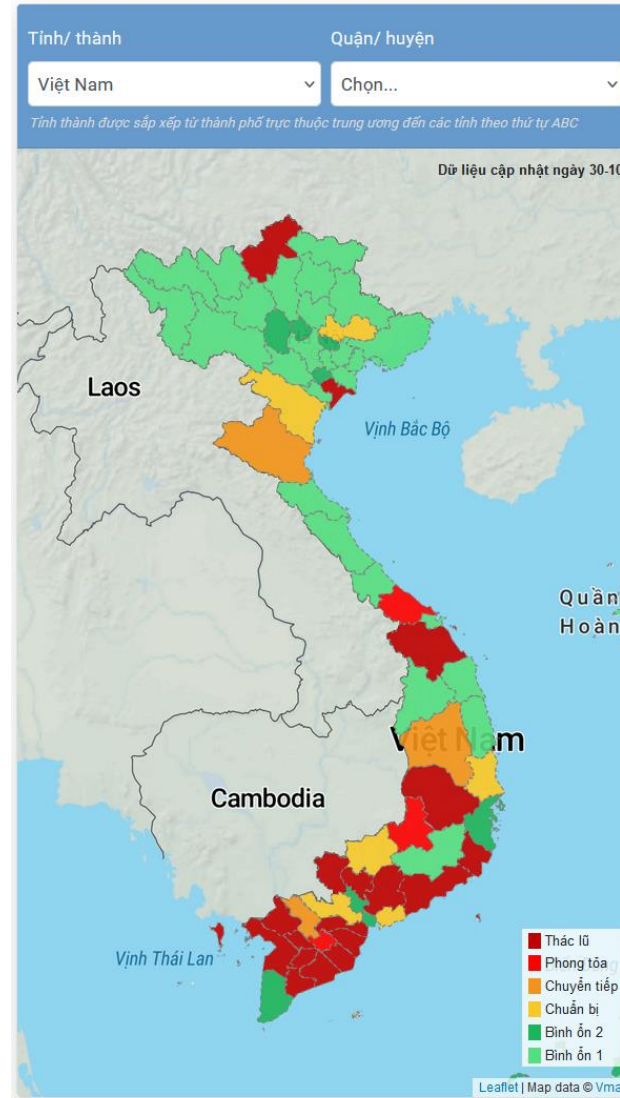
Dự báo % ICU được sử dụng/100k dân					
0-10%	11%-20%	21%-40%	41%-60%	61%-80%	> 80%
A1	A2	B	C	D	E

Thang đo màu sau đó được tùy chỉnh tùy theo: R_T , dự báo số ca nhỏ hay lớn, độ phủ vắc-xin 2 mũi.

Luật di chuyển giữa 2 vùng U → V:

- Khi độ phủ vắc-xin cho mọi vùng là 70% trở lên: dùng màu
- Khi độ phủ vắc-xin của V dưới 70%: dùng mức độ lây nhiễm

Mở rộng cho U → V → W



Kế hoạch trong tương lai

- Cập nhật mô hình khi chúng mới xuất hiện, khi chúng tôi có thêm dữ liệu như số ca và dân số cấp địa lý nhỏ hơn, số ICU mỗi tỉnh.
- Xây dựng các hệ thống tương tự cho các dịch bệnh khác như: bệnh lao, dịch tả lợn châu Phi.
- Hợp tác tích hợp với các hệ thống khác để có hệ thống giám sát dịch tễ hoàn chỉnh.

Trân trọng cảm ơn!

Để biết thêm chi tiết, [website Onyx](#) có loạt 6 bài viết của chúng tôi và nhiều báo cáo khác. Cảm ơn sự đóng góp của:

- Nhóm bạn A0 96-99:
 - Trần Tuấn Anh – chuyên gia Giải pháp và Công nghệ
 - Nguyễn Thị Vân Anh – chuyên gia Chính sách công
 - Nguyễn Mai Nguyệt – chuyên gia Quản trị rủi ro
- PGS. TS. Võ Trí Hảo – Hiệu trưởng Đại học Gia Định
- Nhóm dev Onyx
 - Trần Tiến Trung
 - Trần Đức Thành
- Nhóm nhập dữ liệu: Đàm Thanh Thủy

Tài liệu tham khảo

1. Wikipedia (no date) *Basic reproduction number*. Available at: https://en.wikipedia.org/wiki/Basic_reproduction_number (Accessed: 25 May 2020).
2. Kermack, W. O. and McKendrick, A. G. (1927) 'A contribution to the mathematical theory of epidemics', *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. The Royal Society London, 115(772), pp. 700–721.
3. Fraser, C. (2007) 'Estimating individual and household reproduction numbers in an emerging epidemic', *PLoS ONE*, 2(8).
4. Robert Koch Institut (2020) *Epidemiologisches Bulletin 17-2020: Krankheitsschwere von COVID-19, Nowcasting: Erkrankungsfälle und Reproduktionszahl*. Available at: www.rki.de/epidbull
5. Cori, A. *et al.* (2013) 'A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics', *American Journal of Epidemiology*. Oxford Academic, 178(9), pp. 1505–1512.
6. He, X. *et al.* (2020) 'Temporal dynamics in viral shedding and transmissibility of COVID-19', *Nature Medicine*. Springer Science and Business Media LLC.
7. Driessche, P. van den (2017) 'Reproduction numbers of infectious disease models', *Infectious Disease Modelling*. KeAi Publishing, 2(3), p. 288.