OLTRE IL PICCO

a cura di Michela Baccini, Giulia Cereda, Anna Gottard, Fabrizia Mealli, Cecilia Viscardi

Versione 6 aprile 2020

INTRODUZIONE

L'epidemia da SARS-CoV-2, il nuovo ceppo di coronavirus individuato dalle autorità cinesi agli inizi del gennaio 2020, si sta rapidamente diffondendo in tutti i continenti, mettendo a dura prova la popolazione, il sistema sanitario e l'economia dei paesi colpiti. L'Italia è stato il primo stato europeo raggiunto dall'epidemia e rappresenta tra i più grandi focolai di SARS-CoV-2 fuori dalla Cina. Il numero totale di casi infetti segnalati nel nostro paese è ad oggi vicino le 125'000 unità e i decessi da COVID19 (la malattia respiratoria causata dal nuovo coronavirus) sono più di 15'000.

Dal primo caso di infezione riportato nella zona di Lodi/Codogno il 21 febbraio, la situazione in Italia si è evoluta rapidamente. Ad oggi tutte le regioni sono colpite, sebbene il maggior numero di casi confermati e di decessi sia circoscritto nel nord del paese. L'epidemia è caratterizzata da una forte contagiosità che si traduce in una crescita esponenziale degli infetti. Gli individui colpiti dall'infezione, soprattutto se soggetti a rischio, anziani e/o affetti da più comorbidità, possono andare incontro a gravi complicanze respiratorie, che richiedono il ricovero in terapia intensiva e possono portare alla morte. La pressione sul sistema sanitario è pertanto molto forte, soprattutto nelle zone maggiormente colpite dall'epidemia. Finora non esiste evidenza di farmaci efficaci contro il virus e la disponibilità di un vaccino appare ancora lontana.

In Cina, l'emergenza è stata affrontata con una completa chiusura dell'area di Hubei, dove circa 60 milioni di persone sono state confinate nelle loro abitazioni per alcune settimane. Al momento, sembra che le misure draconiane cinesi siano state efficaci nello stabilizzare l'epidemia. Seguendo l'esempio della Cina, l'Italia ha attuato misure progressive di distanziamento sociale iniziando da piccole aree circoscritte ai primi casi di infezione, fino a un completo lock-down esteso a tutta la nazione, che al momento vede la maggior parte dei suoi settori produttivi fermi, ad eccezione di alcune categorie essenziali. Ai cittadini è vietato uscire di casa salvo casi di necessità o urgenza comprovate. Misure simili sono state attuate progressivamente nella maggior parte dei paesi Europei.

Alla luce della situazione emergenziale, appare evidente la necessità di:

1. monitorare costantemente l'andamento dell'epidemia;
2. valutare l'efficacia delle misure intraprese;
3. produrre previsioni a breve e medio-lungo termine dell'avanzamento dell'epidemia al fine di indirizzare e pianificare azioni nell'immediato futuro.

Il reale processo di diffusione di una malattia infettiva è molto complesso e dipende da innumerevoli fattori. Tuttavia, sotto alcune assunzioni, la dinamica del processo può essere semplificata e riprodotta attraverso modelli matematici. Questi ultimi consentono di ottenere informazioni circa le quantità chiave che regolano l'andamento dell'epidemia, fare previsioni a breve termine (cosa accadrà da qui a pochi giorni?) e simulare scenari plausibili relativi alla futura evoluzione del processo, consentendo

una valutazione comparativa di politiche alternative da adottare per fronteggiare l'epidemia nel medio-lungo periodo.

Tra i suddetti modelli matematici ci sono i modelli compartimentali, come ad esempio il modello SIRD. Tali modelli si fondano sull'assunzione che ad ogni istante, nel corso dell'epidemia, la popolazione sia divisibile in compartimenti, ossia gruppi di individui che si trovano nello stesso stato e che il passaggio dei soggetti da uno stato all'altro segua semplici regole probabilistiche.

Se un problema cruciale di questi modelli è la loro calibrazione, ovvero l'individuazione dei valori dei parametri più verosimili alla luce dei dati osservati [per una descrizione dell'algoritmo di calibrazione che abbiamo utilizzato è in fase di completamento un'appendice metodologica], in questa specifica applicazione, la difficoltà tecnica si unisce alla criticità relativa alla scelta dei dati su cui calibrare il modello. Al momento abbiamo deciso di utilizzare solo il numero giornaliero di decessi COVID19, perché riteniamo che tra i dati disponibili, che includono anche i conteggi dei positivi al tampone e dei guariti, sia quello più attendibile nell'ottica di ricostruire la dinamica effettiva dell'epidemia (si veda anche la discussione in Maltagliati, 2020). Abbiamo inoltre deciso di lavorare a livello regionale e non su dati nazionali o di macro-area: questo perché ci aspettiamo che i parametri che regolano l'epidemia possano essere eterogenei e dipendenti dal contesto locale. Abbiamo iniziato dall'analisi dei dati della regione Toscana considerando una popolazione iniziale di 3'737'000 abitanti.

QUALCHE RISULTATO

COSA SAREBBE SUCCESSO SENZA APPLICARE LE MISURE DI DISTANZIAMENTO SOCIALE

Per prima cosa ci siamo chiesti come si sarebbe evoluta l'epidemia se non fossero state introdotte le misure di distanziamento sociale come da decreto del 10 marzo e successivi. Per semplicità chiameremo questo scenario "scenario di evoluzione naturale o selvaggia dell'epidemia".

Sotto questo scenario, il nostro modello compartimentale ha stimato una letalità, ovvero un rischio di morte per gli infetti, pari all'1%. Questa stima è coerente con le recenti valutazioni dell'Istituto per gli Studi di Politica Internazionale (ISPI) che stima una letalità per l'Italia del 1.14%. Riguardo alla trasmissibilità del virus, abbiamo stimato un parametro $\mathcal{R}_0$ (numero medio di contagi che partono da un individuo infetto in popolazioni numerose, vedi Appendice) pari a 7: in assenza di misure di distanziamento sociale, ogni infetto in media può trasmettere il virus a 7 persone. Si noti che tale valore di $\mathcal{R}_0$ è stimato per il periodo antecedente alle politiche di distanziamento e pertanto ci consente di valutare cosa sarebbe accaduto in uno scenario pessimistico in assenza di provvedimenti.

In Figura 1 riportiamo una descrizione della corrispondente evoluzione dell'epidemia. Sotto lo scenario di evoluzione selvaggia, avremmo osservato un picco dell'epidemia verso i primi di aprile con oltre 2'000'000 di individui infetti circolanti in regione.

Riguardo alle morti attese (Figura 2), l'enorme numero di decessi complessivi previsti dal modello (circa 36'000 morti) non riflette altro che la letalità dell'infezione stimata dal modello (1% sul totale degli infetti, ovvero sulla popolazione regionale, visto che prima o poi, sotto questo scenario semplificato, tutta la popolazione avrebbe contratto il virus!). Il fatto che la curva dei decessi

raggiunga il suo massimo già in estate riflette l'elevata trasmissibilità del virus. E' proprio questa combinazione di assenza di immunità per gli individui nella popolazione e alta trasmissibilità, che

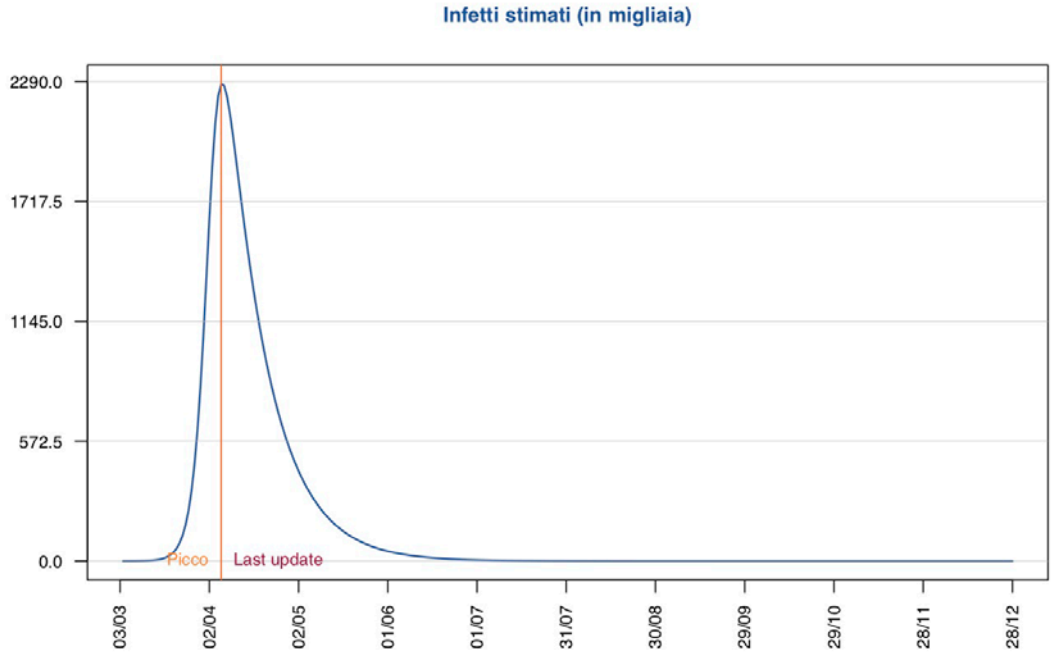
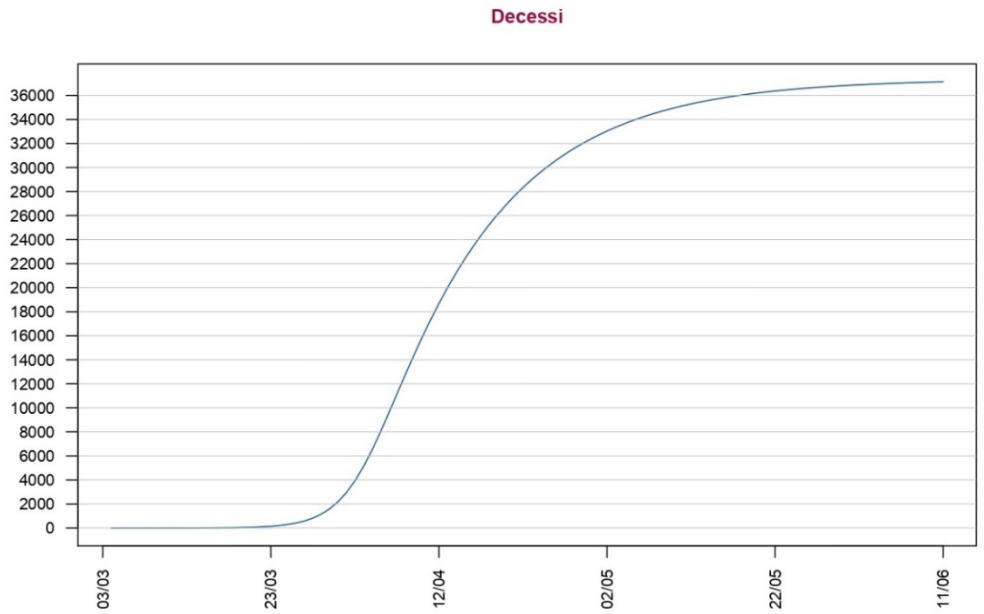


Figura 1 : Scenario in assenza di misure precauzionali – Numero di infetti totale simulato per la Toscana dal 03/03 al 28/12.



rende l'epidemia, se non controllata, particolarmente insidiosa, nonostante la letalità possa essere di per sé non molto elevata. Notiamo anche che l'epidemia si sarebbe probabilmente esaurita prima dell'estate, ma ad un prezzo elevatissimo per la popolazione e il servizio sanitario.

Figura 2: Scenario in assenza di misure precauzionali - Numero di decessi totale simulato per la Toscana dal 03/03 al 11/06.

La Figura 3, che dà il dettaglio sui decessi fino a inizio aprile, mostra che le morti effettivamente osservate (in rosso) sono in accordo al modello di evoluzione selvaggia fino al 21-22 marzo, per poi

discostarsi nei giorni successivi, ad indicare che in Toscana le politiche di contenimento stanno avendo un effetto.

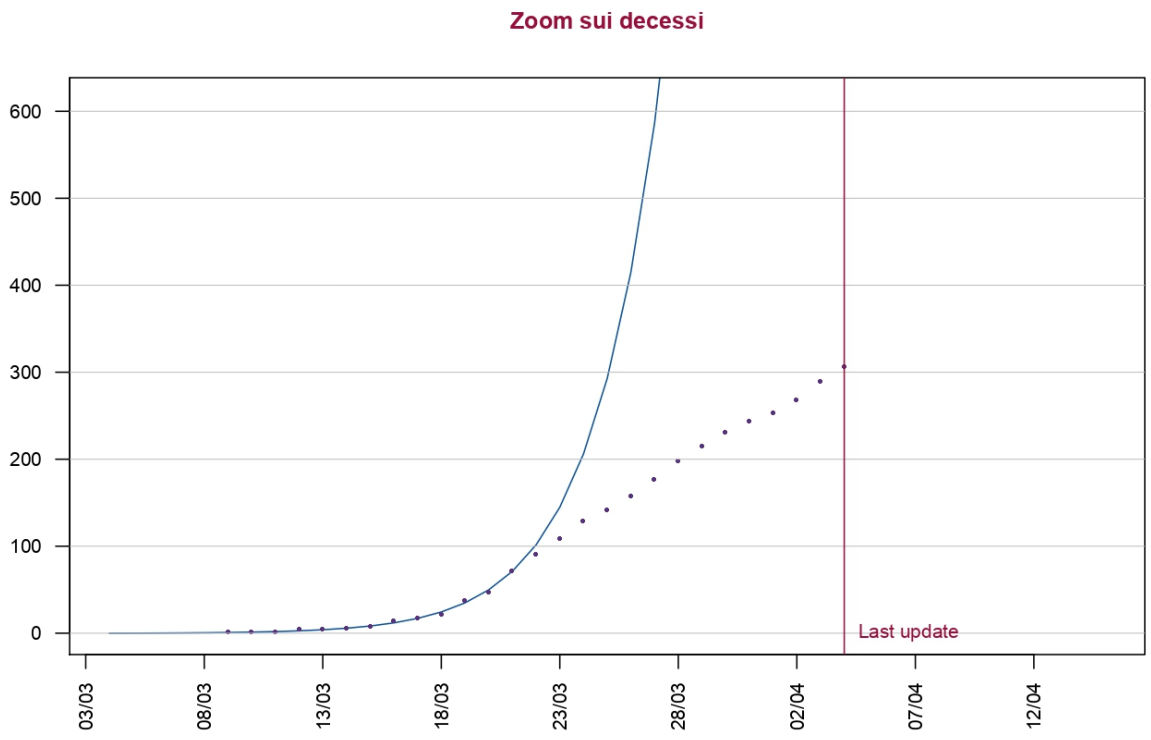


Figura 3: Scenario in assenza di misure precauzionali - Zoom del numero di decessi totale simulato per la Toscana fino al 05/04.Curva: numero di decessi simulati; Punti: numero di decessi osservati.

COSA POSSIAMO ASPETTARCI SE RESTIAMO A CASA

Il nostro secondo obiettivo è quello di valutare quale potrebbe essere l'evoluzione dell'epidemia sotto lo scenario attuale di lock-down come implementato in regione. E' necessaria una premessa: le misure di distanziamento sociale fronteggiano la diffusione del virus riducendo $\mathcal{R}_0$. Esse non incidono sul livello di letalità dell'infezione, né sui tempi di guarigione o morte dei malati. Pertanto, quello che ci proponiamo di fare è valutare progressivamente, sulla base delle informazioni aggiornate sul numero di decessi, quanto ci si stia discostando dall'$\mathcal{R}_0$ dello scenario di evoluzione selvaggia, fermi restando gli altri parametri chiave che descrivono l'epidemia.

Siamo partiti dall'ipotesi che se "stiamo a casa" $\mathcal{R}_0$ decresca: un andamento possibile, che al momento sembra adattarsi bene alla mortalità osservata, è quello che vede una progressiva diminuzione di $\mathcal{R}_0$ da 7 a 0.50 nell'arco di 24 giorni dal 10 marzo, secondo la curva riportata in Figura 4. Raggiungere un $\mathcal{R}_0$ minore di 1 è fondamentale per far sì che l'epidemia si esaurisca senza che il virus si diffonda a tutta la popolazione. Questa ipotesi di decremento di $\mathcal{R}_0$, che, come già detto, ad oggi sembra essere coerente con la serie dei decessi osservati (si veda la Figura 7 che fornisce il dettaglio sui decessi fino al 5 aprile), potrà o meno essere confermata dai dati di mortalità che saranno rilevati nei prossimi giorni.

Tutto ciò premesso, come mostrato in Figura 5, il nostro modello prevede che si sia raggiunto il picco dell'epidemia intorno al 28 marzo, con un numero di soggetti affetti da coronavirus pari a circa

28'000, molti meno di quelli che ci saremmo dovuti aspettare sotto lo scenario di evoluzione selvaggia. Dopo il picco, si assisterà ad un progressivo calo del numero di infetti. Dobbiamo immaginare che, tanto più basso sarà il valore di $\mathcal{R}_0$ che riusciremo a raggiungere - rispetto allo 0.50 ipotizzato dall'attuale modello - tanto più veloce sarà il declino della curva. Sempre secondo il nostro modello, anche il numero dei morti sarà contenuto rispetto alla stima selvaggia: complessivamente i decessi potrebbero essere intorno a 700 decessi a metà giugno (Figura 6).

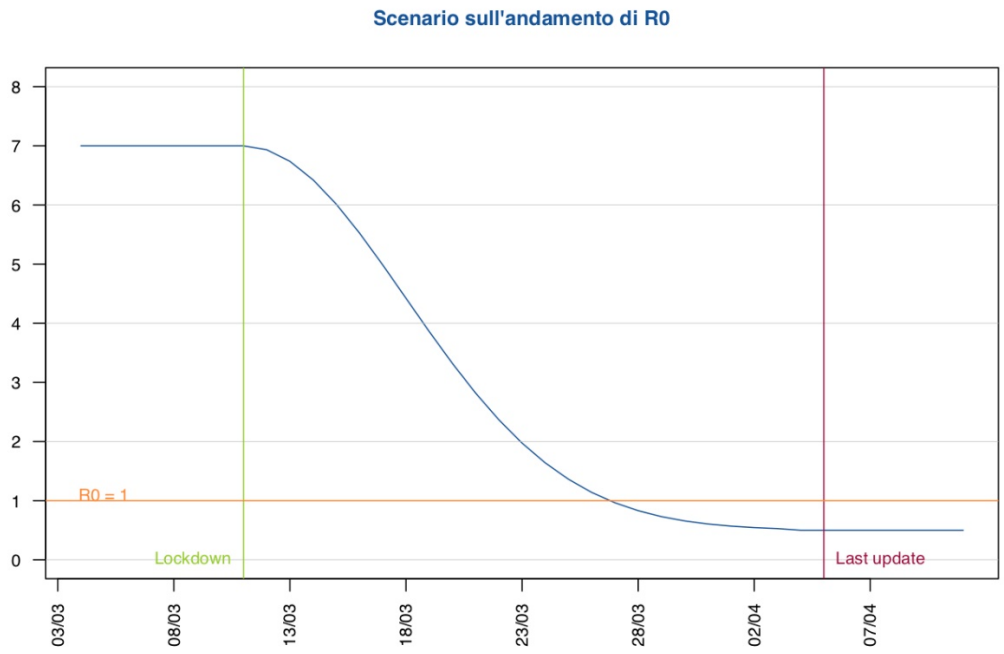


Figura 4: Evoluzione di $\mathcal{R}_0$ - $\mathcal{R}_0$ costante fino al 10/03 (primo decreto Conte per la Toscana) e decrescente dall'11/03.

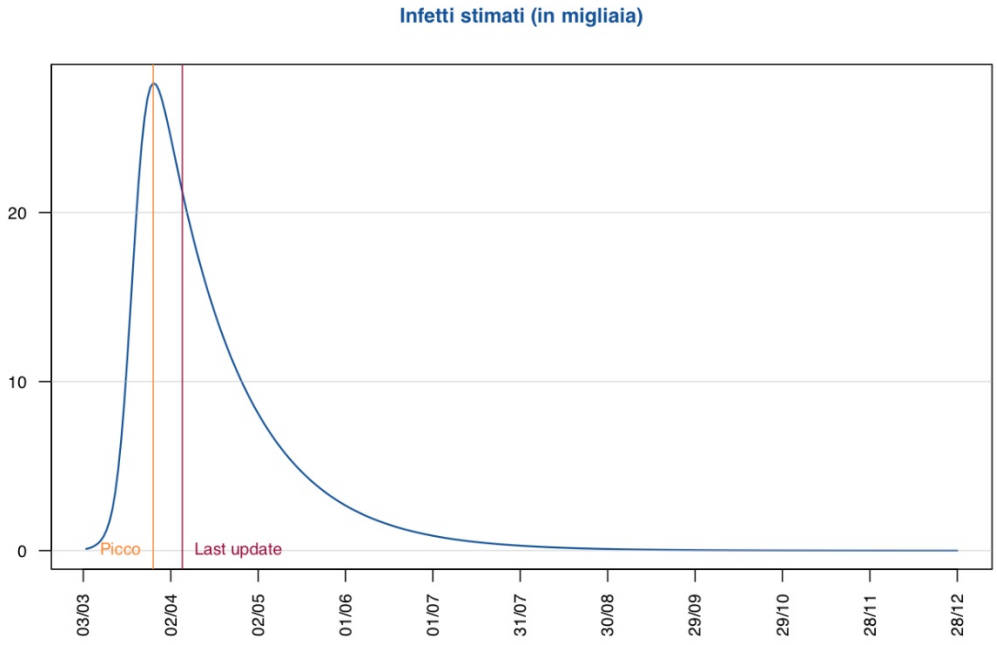


Figura 5: Scenario con politiche di contenimento - Numero di infetti totale simulato per la Toscana dal 04/03 al 28/12.

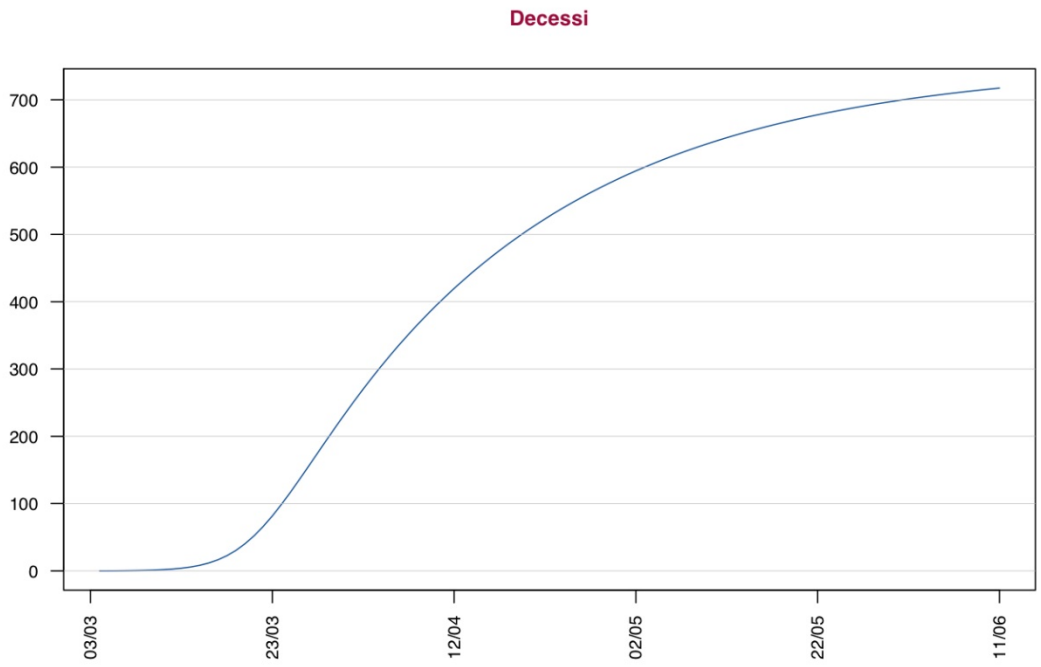


Figura 6: Scenario con politiche di contenimento - Numero di decessi totale simulato per la Toscana dal 04/03 all'11/06.

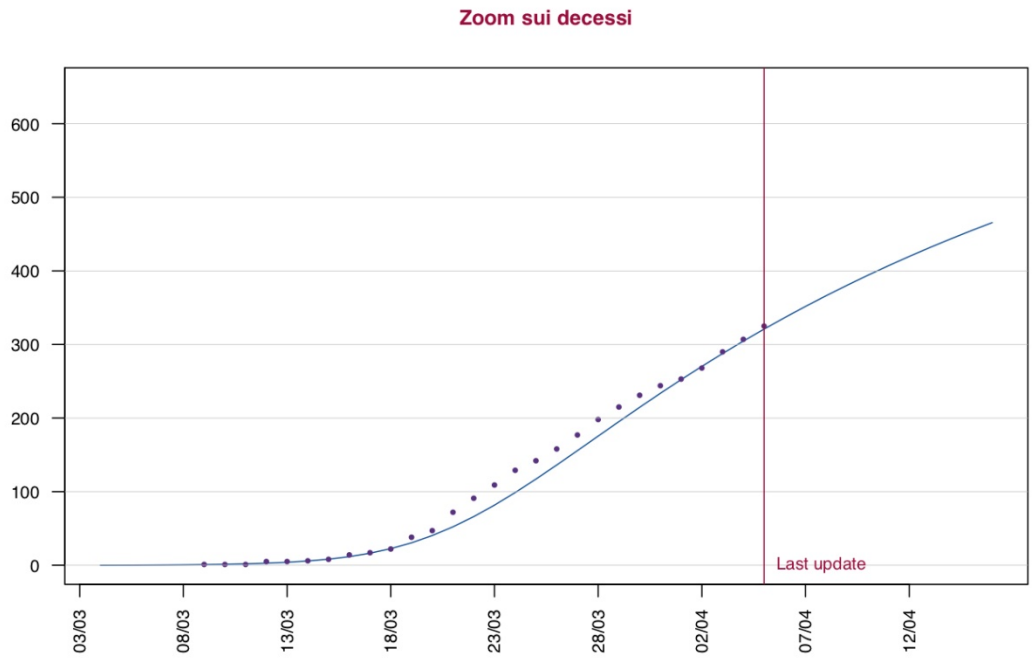


Figura 7: Scenario con politiche di contenimento - Zoom del numero di decessi totale simulato per la Toscana fino al 05/04.

Curva: numero di decessi simulati; Punti: numero di decessi osservati.

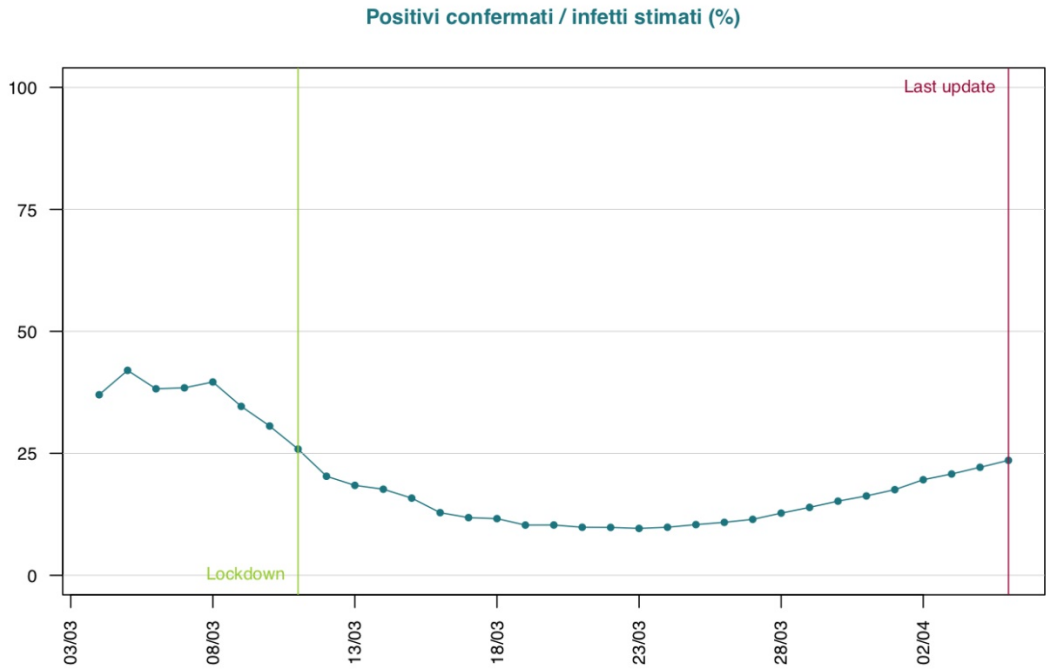


Figura 8: Scenario con politiche di contenimento - Rapporto tra numero osservato di tamponi positivi e numero di infetti previsti dal modello.

In Figura 8 abbiamo riportato le stime giornaliere della percentuale di positivi al tampone riportati dalle statistiche ufficiali (il numero di positivi dall'inizio dell'epidemia meno il numero dei decessi e dei guariti) sul numero totale di infetti predetto dal modello. I nostri calcoli sembrano indicare una progressiva diminuzione di questa percentuale, che, da valori all'inizio del mese di marzo compresi tra il 25% e il 40%, cala fino al 10%. Attualmente sembra assestarsi intorno al 25%. Questo significa che ad oggi è possibile che per ogni positivo registrato ci siano 3 infetti non individuati.

Un'ultima ma essenziale considerazione riguarda la nostra ipotesi sull'evoluzione di $\mathcal{R}_0$. Nel nostro modello abbiamo assunto che da fine mese $\mathcal{R}_0$ resterà costante (a 0.50); questo equivale a ipotizzare che il lock-down continuerà per un tempo estremamente lungo (ipotesi economicamente e socialmente insostenibile) o che sarà sostituito da misure altrettanto efficaci. Nella migliore delle ipotesi, l'estate potrà essere di aiuto, ma non sono ancora chiare le evidenze riguardo alla circolazione del virus nella stagione calda. E' pertanto necessario sviluppare quanto prima un piano di azione ragionato che combini strategie di distanziamento sociale o isolamento parziale con metodi di individuazione e isolamento precoce dei positivi, soprattutto nei contesti dove il rischio di diffusione è maggiore (comunità, ospedali). Tutto questo nel rispetto delle libertà fondamentali dell'individuo.

CARATTERISTICHE E UTILIZZO DEL NOSTRO APPROCCIO

Il nostro approccio, che per molti aspetti è simile a quello proposto da altri gruppi che come noi in questi giorni hanno lavorato in modo concitato nell'interesse comune, ha delle caratteristiche peculiari che qui elenchiamo brevemente:

- Rispetto ai modelli che si concentrano su dati a livello nazionale o di macro-area, l'utilizzo di dati regionali consente di ridurre l'eterogeneità dei parametri che regolano l'epidemia,

ottenendo stime e previsioni maggiormente aderenti al contesto locale, quindi potenzialmente più affidabili;

- Abbiamo prodotto uno scenario controfattuale di evoluzione selvaggia dell'epidemia, rispetto al quale potranno essere valutate anche politiche di contenimento diverse da quella effettivamente intrapresa, per esempio misure di distanziamento sociale più blande o concentrate su specifici sottogruppi della popolazione;
- Il nostro approccio, pur in un contesto di estrema semplificazione di un fenomeno complesso, dovrebbe consentire il monitoraggio dell'effetto delle politiche intraprese (seguiremo l'evoluzione di $\mathcal{R}_0$ nel tempo) e fare previsioni sull'impatto di politiche future, sia a breve che a lungo termine. Per esempio, il modello può essere utilizzato per simulare scenari alternativi di graduale rilassamento delle misure di distanziamento sociale;
- La definizione del nostro modello si basa su parametri che hanno un significato epidemiologico chiaro e semplice. Il modello può quindi essere integrato con valutazioni di esperti o con evidenze derivanti da altri studi sull'argomento.

LIMITAZIONI

L'approccio che abbiamo adottato ha diverse limitazioni che riassumiamo in tre punti:

- Il modello SIRD che abbiamo usato si basa su ipotesi forti, riducendo a semplici equazioni l'evoluzione temporale di un fenomeno molto complesso. Per esempio, il modello non tiene conto che la letalità è diversa per classi di età e per livello di fragilità della popolazione, che $\mathcal{R}_0$ può variare molto a seconda del contesto, che possono esserci cluster di epidemia e che, anche entro la regione, l'evoluzione dell'infezione può variare geograficamente (banalmente, l'inizio dell'epidemia potrebbe essere diverso a seconda dell'area considerata);
- In accordo al nostro modello, i parametri "epidemiologici" che regolano l'evoluzione dell'epidemia sono: $\mathcal{R}_0$, letalità, tempo dall'infezione all'eventuale morte del soggetto, tempo dall'infezione all'eventuale guarigione, data di inizio epidemia. Dato che combinazioni diverse di questi parametri possono portare a stesse previsioni (una situazione che in statistica è chiamata non identificabilità), abbiamo per semplicità fissato a 14 giorni il tempo di guarigione e di morte. Una scelta diversa di questi tempi potrebbe portare a stime differenti dei restanti parametri (precisiamo comunque che al momento la scelta di 14 giorni è quella che appare "migliore" in termini di adattamento ai dati di mortalità osservati, rispetto a 21 e 28 giorni);
- Con il nostro approccio cerchiamo quale combinazione di parametri meglio riproduce i decessi osservati, ovvero facciamo una cosiddetta "calibrazione" sul numero dei decessi. Per esempio, nel modello usato per studiare l'evoluzione selvaggia dell'epidemia, su un insieme grande di possibili combinazioni di $\mathcal{R}_0$, letalità e data di inizio epidemia, abbiamo selezionato quella che produceva un numero di decessi, da inizio epidemia a 10 giorni dall'introduzione del lock-down, più simile possibile all'osservato. Combinazioni alternative potrebbero tuttavia produrre risultati simili. Inoltre, ci potrebbero essere, tra quelle sperimentate, combinazioni di parametri a priori più plausibili di altre, anche alla luce delle evidenze che giorno dopo giorno arrivano da studi condotti sull'epidemia; il nostro modello al momento non tiene conto di questo aspetto (anche se ci stiamo attrezzando, si veda la sezione WIP).

WORK IN PROGRESS (WIP)

Al momento, stiamo lavorando sui seguenti punti:

- Previsioni a medio/lungo termine;
- Analisi dei dati per tutte le regioni italiane, eventualmente anche a livello provinciale;
- Propagazione dell'incertezza dai parametri di input del modello ($\mathcal{R}_0$, letalità, tempi di guarigione e morte) alle previsioni.
- Definizione di distribuzioni di probabilità a priori sui parametri del modello e aggiornamento degli stessi alla luce dei dati tramite tecniche Bayesiane;
- Confronto con modelli utilizzati da altri ricercatori e risultati ottenuti in contesti geografici diversi.

Appendice: Descrizione di un modello compartimentale tipo SIRD - Draft

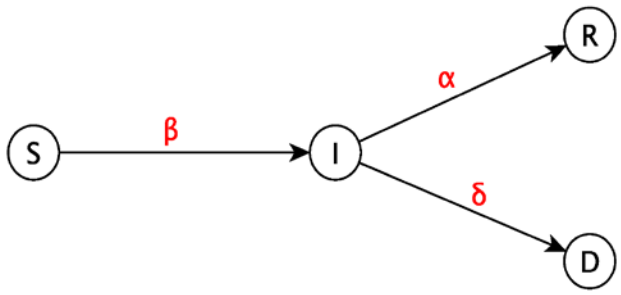
Il modello che abbiamo adottato è un modello compartimentale di tipo SIRD (Susceptible-Infective-Recovered-Dead) in cui la popolazione è suddivisa, ad ogni istante temporale, nelle seguenti quattro classi:

1. Suscettibili: individui che possono contrarre la malattia;
2. Infetti: individui che sono stati contagiati e che possono contagiare;
3. Guariti;
4. Morti.

Con $S(t), I(t), R(t)$ e $D(t)$ sono indicati rispettivamente il numero di individui in ciascuno dei quattro compartimenti nell'istante temporale t.

La numerosità di ciascun gruppo varia nel corso del tempo e descrive la dinamica di diffusione dell'infezione. Assumendo che il tempo sia discreto (nelle nostre analisi l'unità temporale è il giorno), la dimensione dei gruppi al tempo t dipende solo da quanto accaduto al tempo $t-1$ secondo le seguenti equazioni:

1. $I(t) = I(t-1) + \beta S(t-1)I(t-1) - \alpha I(t-1) - \delta I(t-1)$
2. $R(t) = R(t-1) + \alpha I(t-1)$
3. $D(t) = D(t-1) + \delta I(t-1)$



Parametri

Ciascun individuo nella popolazione può transitare da uno stato ad un altro e le transizioni possibili sono rappresentate dalle frecce riportate nel seguente grafico.

Tali transizioni sono regolate dai seguenti tassi:

- Tasso di trasmissione (β):

  Indicando con β_0 la probabilità che un contatto tra due individui porti ad un contagio e con $\bar{c}$ il numero medio di contatti di un individuo in una unità di tempo, il tasso di trasmissione può essere espresso come $\beta = \frac{\beta_0 \cdot \bar{c}}{N}$, dove N indica la numerosità della popolazione.

- Tasso di guarigione (α):

  Guardando alla figura si vede che la malattia ha due soli possibili esiti, guarigione o morte. Assumendo che α_0 sia la probabilità che ciascun individuo infetto ha di guarire e che T_r

  sia il tempo medio di guarigione, il tasso di guarigione può essere espresso come $\alpha = \alpha_0/T_r$

- Tasso di letalità (δ):

  Il secondo possibile esito della malattia è la morte, per cui $1 - \alpha_0$ rappresenta la probabilità che ciascun individuo infetto ha di morire. Ne consegue che, assumendo che il tempo medio che intercorre tra il contagio e la morte sia T_d, il tasso di letalità può essere espresso come

  $\delta = \frac{1-\alpha_0}{T_d}$.

  Oltre ai tassi di transizione, un parametro cardine in epidemiologia è il *tasso netto di riproduzione*, $\mathcal{R}_0$. Quest'ultimo rappresenta il numero medio di contagi generati dal primo individuo infetto all'inizio dell'epidemia, ossia quando l'intera popolazione è suscettibile.

  Al fine di quantificare la capacità di propagazione dell'epidemia è fondamentale monitorare $\mathcal{R}_0$ che, in popolazioni sufficientemente numerose, può essere interpretato come il numero di nuovi infetti che si sostituiscono ad un individuo che guarisce o muore. Questa interpretazione consente di comprendere che la diffusione di un'epidemia può arrestarsi solo quando $\mathcal{R}_0 < 1$. In tal caso, infatti, il numero di individui che escono dal compartimento degli infetti è sempre maggiore di quelli che vi entrano. Viceversa, $\mathcal{R}_0 > 1$ implica che per ciascun individuo che guarisce o muore si ha più di un nuovo contagio.

Indicando con $\Delta I, \Delta S, \Delta R$ e ΔD rispettivamente la variazione della numerosità dei rispettivi compartimenti tra i tempi $t = 0$ e $t = 1$, $\mathcal{R}_0$ può essere calcolato come segue:

$$\mathcal{R}_0 = \frac{\Delta I + \Delta R + \Delta D}{\Delta R + \Delta D} = \frac{\beta I(0)S(0) - \alpha I(0) - \delta I(0) + \alpha I(0) + \delta I(0)}{\alpha I(0) + \delta I(0)}$$

da cui, essendo $I(0) = 1$

$$\mathcal{R}_0 = \frac{\beta S(0)}{\alpha + \delta}$$

Questa uguaglianza stabilisce la seguente relazione tra il tasso di trasmissione e $\mathcal{R}_0$:

$$\beta = \frac{\mathcal{R}_0\,(\alpha+\delta)}{S(0)}.$$

Assunzioni del modello

1. L'epidemia ha inizio in una popolazione composta da tutti suscettibili ed un solo individuo già infetto;
2. La popolazione è chiusa: non ci sono migrazioni, nascite e morti (ad eccezione di quelle di individui affetti da Covid-19). Dunque, sia N la numerosità iniziale della popolazione, ad ogni istante t vale la seguente uguaglianza:

$$S(t) + I(t) + R(t) + D(t) = N$$

3. Ciascun individuo appartenente alla classe dei suscettibili ha la stessa probabilità di essere contagiato;
4. Una volta contratto il virus la probabilità di tornare ad essere suscettibile è nulla (fissiamo il tasso di perdita dell'immunità come pari a 0);
5. La probabilità di contagiare un suscettibile è uguale per tutti gli infetti. Lo stesso vale per la probabilità di guarire e di morire;
6. I due soli possibili esiti della malattia sono la guarigione e la morte;
7. Gli individui appartenenti alle classi dei guariti e dei morti non sono più infettivi;
8. Ciascun individuo contagiato inizia ad essere infettivo dall'istante successivo al contagio.