

Analisi e previsioni dell'epidemia da SARS-CoV-2 in Toscana

Analysis and future scenarios of the SARS-CoV-2 epidemic in Tuscany Region (Central Italy)

Giulia Cereda,^{1,2} Cecilia Viscardi,^{1,2} Anna Gottard,^{1,2} Fabrizia Mealli,^{1,2} Michela Baccini^{1,2}

¹ Dipartimento di statistica, informatica, applicazioni (DISIA), Università di Firenze

² Florence Center for Data Scienze, Università di Firenze

Corrispondenza: Michela Baccini; michela.baccini@unifi.it

RIASSUNTO

OBIETTIVI: a circa due mesi dalla fine del *lockdown* imposto per il contenimento dell'epidemia di SARS-CoV-2, si sono valutate le dinamiche del contagio nella Regione Toscana dall'inizio dell'emergenza a fine giugno attraverso un modello compartimentale, successivamente utilizzato per produrre semplici previsioni degli andamenti epidemici per i prossimi mesi.

DATI E METODI: è stato costruito un modello compartimentale di tipo SIRD, in cui il tasso di riproduzione dell'infezione (R_0) è stato assunto variabile nel tempo, tramite una funzione costante a tratti. Fissati, per motivi di identificabilità, alcuni parametri – la letalità e il tempo dal contagio alla risoluzione dell'infezione (morte o guarigione) – il modello è stato calibrato sulla serie dei decessi per malattia Covid-19 notificati nel periodo dal 09.03.2020 al 30.06.2020. L'incertezza attorno alle stime è stata quantificata attraverso un *bootstrap* parametrico. Il modello stimato è stato poi utilizzato per produrre proiezioni a medio-lungo termine delle dinamiche epidemiche.

RISULTATI: la data di inizio epidemia in Toscana è stata stimata al 21.02.2020. Il valore stimato di $R_0(t)$ è andato da un iniziale 7,78 (IC95% 7,55-7,89) a un valore molto vicino a 0 tra il 27 aprile e il 17 maggio. Infine, esso è risalito, fino a raggiungere un valore medio di 0,66 (0,32-0,88) tra il 18 maggio e il 30 giugno. In corrispondenza del picco epidemico, stimato all'inizio di aprile, gli infetti notificati circolanti in regione risultavano appena il 22% di quelli predetti dal SIRD. In accordo al modello, se a partire da ottobre $R_0(t)$ superasse, anche di poco, l'unità, l'onda del contagio potrebbe raggiungere nuovamente livelli preoccupanti entro la prossima primavera.

CONCLUSIONI: l'andamento stimato di $R_0(t)$ suggerisce la presenza di un forte effetto delle politiche di contenimento sulla diffusione del virus in Toscana e di una minima ripresa del contagio potenzialmente attribuibile all'interruzione del *lockdown*. Le proiezioni a medio-lungo termine mostrano inequivocabilmente come il pericolo di una nuova ondata epidemica non sia scongiurato.

Parole chiave: modelli compartimentali, numero di riproduzione dell'infezione, SIRD, previsioni, SARS-CoV-2

ABSTRACT

OBJECTIVES: about two months after the end of the lockdown imposed for the containment of the SARS-CoV-2 epidemic, the contagion dynamics in the Tuscany Region (Central Italy) have been assessed from the beginning of the emergency to the end of June through a compartmental model, and future medium-long term projections have been produced.

DATA AND METHODS: this study used a SIRD model in which the infection reproduction number R_0 varied over time, according to a piecewise constant function. The fatality pa-

COSA SI SAPEVA GIÀ

■ È importante modellare le dinamiche di contagio da SARS-CoV-2 sia per valutare retrospettivamente l'efficacia delle misure intraprese per il contenimento dell'epidemia sia per fare previsione a medio-lungo termine.

■ I modelli compartimentali di tipo SIRD consentono di modellare l'andamento dell'epidemia nel tempo, cioè la transizione dei soggetti dallo stato di suscettibili allo stato di infetti e dallo stato di infetti a quello di guariti o deceduti.

■ L'implementazione di forti misure di contenimento e altri fattori di tipo ambientale può influire sulle dinamiche epidemiche e in particolare sul numero di riproduzione dell'infezione ($R_0(t)$), che può variare nel tempo.

COSA SI AGGIUNGE DI NUOVO

■ In Toscana, il valore di R_0 all'inizio dell'epidemia (valore pre-*lockdown*) era maggiore di 7: mediamente ogni soggetto infetto era capace di contagiare altri 7,78 individui.

■ Dopo il *lockdown* c'è stata una forte e immediata riduzione dell'indice R_0 e, alla riapertura, un lieve aumento di questo indice. Questo suggerisce che le misure di contenimento hanno verosimilmente avuto una forte influenza sulle dinamiche di contagio.

■ Se il valore di $R_0(t)$ a ottobre dovesse risalire, anche di poco, oltre la soglia di 1, si potrebbe assistere a una nuova ondata di contagi entro la primavera.

rameter and the time from contagion to infection resolution (death or recovery) were fixed to ensure parameter identifiability, and the model was calibrated on the Covid-19 deaths notified from March 9th to June 30th 2020. The uncertainty around the estimates was quantified through a parametric bootstrap. Finally, the resulting model was used to produce medium-long term projections of the epidemic dynamics.

RESULTS: the date of the first infection in Tuscany was estimated as February 21st 2020. The value of $R_0(t)$ ranged from 7.78 (95%CI 7.55-7.89), at the beginning of the outbreak, to a value very close to 0 between April 27th and May 17th. Finally, $R_0(t)$ rose, reaching an average of 0.66 (0.32, 0.88) between May 18th and June 30th. At the epidemic peak, estimated at the beginning of April, the notified infected people circulating in the region were just 22% of those predicted by the model. According to the estimated SIRD, under the hypothetical scenario that $R_0(t)$ slightly exceeds 1 from the beginning of October 2020, a new wave of contagion could arise by next spring.

CONCLUSIONS: the estimated trend of $R_0(t)$ is suggestive of a strong effect of the lockdown in Tuscany and of a mild increase of the contagion potentially attributable to the easing of the containment measures. Medium-long term projections unequivocally indicate that the danger of a new epidemic wave has not been averted.

Keywords: compartmental models, basic reproduction number, SIRD, projections, SARS-CoV-2

INTRODUZIONE

A partire dall'inizio di gennaio 2020, l'epidemia da SARS-CoV-2 si è diffusa rapidamente dalla Cina a tutti i continenti, mettendo a dura prova la popolazione, il sistema sanitario e l'economia dei paesi colpiti. L'Italia è stato uno dei più grandi focolai di SARS-CoV-2 in Europa, nonché il primo Paese del continente ad essere colpito dalla pandemia. In totale, a fine giugno si sono contati circa 240.000 casi notificati e circa 35.000 decessi da Covid-19 (la malattia respiratoria causata dal nuovo coronavirus), specialmente tra soggetti anziani e/o affetti da più comorbidità.

Dopo mesi di forte pressione sul sistema sanitario, a oggi (luglio 2020) la situazione sembra essere sotto controllo, fatta eccezione per la Lombardia e alcuni focolai circoscritti, anche a seguito delle progressive misure di distanziamento interpersonale, risultate in un periodo di *lockdown* di circa due mesi, terminato all'inizio del mese di maggio. Il numero di nuovi casi è contenuto e molte regioni non presentano nuovi contagi. Attualmente l'Italia si trova nella cosiddetta «fase due»: le attività sono riprese con l'obbligo di rimanere distanziati nei luoghi chiusi e seguire regole che disciplinano la sanificazione, l'utilizzo di dispositivi sanitari quali le mascherine e i gel disinfettanti. Le scuole e le università, invece, non hanno ripreso l'attività didattica in presenza.

A circa due mesi dalla riapertura, appare opportuno valutare le dinamiche epidemiche dall'inizio dell'emergenza, con l'intento di valutare retrospettivamente l'efficacia delle misure intraprese per il contenimento del contagio e produrre scenari futuri a medio-lungo termine del possibile avanzamento dell'epidemia che, per quanto ipotetici, possano indirizzare e pianificare azioni nell'immediato futuro. Inoltre, la messa a punto di strumenti modellistici per lo studio delle dinamiche epidemiche potrà rivelarsi uno strumento utile per la sorveglianza nei mesi a venire.

La dinamica di diffusione di un'epidemia è molto complessa, ma, sotto alcune assunzioni, può essere semplificata e riprodotta attraverso modelli matematici. Tra questi, i modelli di tipo compartimentale si fondano sull'assunzione che a ogni istante nel corso dell'epidemia la popolazione possa essere suddivisa in compartimenti, ossia gruppi di individui che si trovano nello stesso stato rispetto al fenomeno in studio e che il passaggio dei soggetti da uno stato all'altro segua semplici regole probabilistiche.¹ Esistono molti tipi di modelli compartimentali, definiti di solito sulla base dei compartimenti specificati. Per fare due esempi, il modello SIR include tre compartimenti: suscettibili (Susceptible), infetti (Infected) e guariti o deceduti (Removed); il modello SIRD divide il compartimento dei rimossi in guariti (Recovered) e deceduti (Dead).

In questo articolo presentiamo i risultati di un'analisi in cui è stato utilizzato un modello compartimentale di tipo SIRD,² calibrato sui decessi da Covid-19 notificati, per studiare la diffusione del contagio da SARS-CoV-2 nel-

la regione Toscana. L'obiettivo di questo studio è duplice: ottenere informazioni circa una delle quantità chiave che regolano l'andamento dell'epidemia, il tasso di riproduzione dell'infezione, e fare previsioni a lungo termine simulando scenari plausibili relativi alla futura evoluzione del contagio.

DATI

L'analisi si è basata sui dati giornalieri messi a disposizione dalla Protezione civile e in particolare sul numero giornaliero di decessi Covid-19 notificati dal 09.03.2020, giorno del primo decesso in Regione, al 30.06.2020. Tra le informazioni rese disponibili dalla Protezione civile vi sono anche i conteggi dei positivi al tampone, dei guariti, degli individui in isolamento domiciliare e degli ospedalizzati. Il dato di mortalità, tuttavia, è apparso, da una parte facilmente utilizzabile per calibrare un modello compartimentale SIRD – che include il compartimento dei deceduti –, dall'altra il più attendibile nell'ottica di ricostruire la dinamica effettiva dell'epidemia.³

METODI

Il modello compartimentale di tipo SIRD (*Susceptible-Infective-Recovered-Dead*) divide la popolazione a ogni istante temporale, nei seguenti compartimenti:²

- **Suscettibili (S):** individui che possono contrarre la malattia;
- **Infetti (I):** individui che sono stati contagiati e che possono contagiare;
- **Guariti (R):** soggetti che hanno contratto la malattia e sono guariti;
- **Morti (D):** soggetti che hanno contratto la malattia e sono morti.

Le frecce in figura 1 rappresentano le transizioni possibili tra i vari compartimenti, che sono governate dalle seguenti equazioni:

$$\begin{aligned} S(t+1) &= S(t) - \beta S(t) I(t) / S(0) \\ I(t+1) &= I(t) + \beta S(t) I(t) / S(0) - \alpha I(t) - \delta I(t) \\ R(t+1) &= R(t) + \alpha I(t) \\ D(t+1) &= D(t) + \delta I(t) \end{aligned}$$

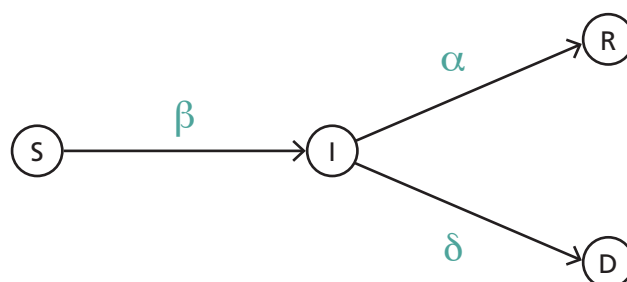


Figura 1. Rappresentazione grafica delle transizioni possibili tra i compartimenti del modello SIRD.

Figure 1. Allowed transitions across the compartments of the SIRD model.

dove $S(t)$, $I(t)$, $R(t)$ e $D(t)$ rappresentano la numerosità di ciascuno dei quattro compartimenti nell'istante temporale t , assumendo una modellazione in tempo discreto in cui l'unità temporale è il giorno. All'inizio dell'epidemia ($t = 0$) si ha: $I(0) = 1$, $S(0) = N - 1$, $R(0) = D(0) = 0$, dove N è la numerosità iniziale della popolazione, in questo caso $N = 3.737.000$. Si noti che la dimensione dei gruppi al tempo $t + 1$ dipende solo dalla loro dimensione al tempo t e che la somma totale degli individui a ogni istante è costante e pari a $S(t) + I(t) + R(t) + D(t) = N$.

Le transizioni da un compartimento all'altro sono regolate dai seguenti parametri: β , da cui dipende il tasso di trasmissione del contagio e quindi il numero di nuovi soggetti infetti a ogni istante di tempo; il tasso di letalità δ ; il tasso di guarigione α . δ e α possono essere così espressi:

$$\delta = p^* / T_d$$

$$\alpha = (1 - p^*) / T_r$$

dove T_d è il tempo medio che intercorre tra il contagio e la morte, T_r il tempo medio tra il contagio e la guarigione e p^* è la proporzione di individui destinati a transitare verso il compartimento dei deceduti tra quelli che sono presenti nel compartimento degli infetti in uno stesso istante temporale. La proporzione p^* è, pertanto, calcolata su un insieme di soggetti che sono stati infettati in momenti diversi e, tenendo conto che questi sono una mistura di individui che guariscono o muoiono con tempi differenti, si differenzia dalla probabilità di morte per gli infetti o letalità, ed è a essa legata dalla relazione:⁴

$$p^* = \frac{p T_d}{p T_d + (1 - p) T_r}$$

Ne consegue che, se $T_d = T_r$, i parametri p e p^* sono equivalenti. Si noti che $p = \delta / (\alpha + \delta)$.

Un parametro cardine del modello è il tasso effettivo di riproduzione, R_t . Quest'ultimo rappresenta il numero medio di contagi generati da un individuo infetto al tempo t . Il suo legame con i parametri α , δ e β è facilmente ricavabile: indicando con ΔI , ΔS , ΔR , e ΔD , la variazione della numerosità dei rispettivi compartimenti tra i tempi t e $t + 1$, R_t può essere calcolato come segue:

$$R_t = \frac{\Delta I + \Delta R + \Delta D}{\Delta R + \Delta D} = \frac{\beta I(t) S(t) / S(0) - \alpha I(t) - \delta I(t) + \alpha I(t) + \delta I(t)}{\alpha I(t) + \delta I(t)} = \frac{\beta}{\alpha + \delta} \frac{S(t)}{S(0)}$$

In particolare, se $t = 0$ si ottiene il *tasso netto di riproduzione*, $R_0 = \beta / (\alpha + \delta)$. Dunque, β può essere espresso in funzione di R_0 tramite:

$$\beta = R_0 (\alpha + \delta)$$

Si noti che tra il *tasso effettivo di riproduzione* e il *tasso netto di riproduzione* vale la seguente relazione:

$$R_t = R_0 \frac{S(t)}{S(0)}$$

R_t decresce a mano a mano che il numero dei suscettibili diminuisce, cioè quando gli infetti muoiono o, guarendo, diventano immuni. Ovviamente se, come in questo caso, la numerosità del compartimento S è molto elevata e si è in uno stadio iniziale dell'epidemia, il rapporto $S(t)/S(0)$ è approssimabile a 1 e $R_t \approx R_0$.

Nonostante nel SIRD classico R_0 sia costante e riferito al tempo $t = 0$, è possibile specificare una versione modificata del modello in cui R_0 vari nel corso dell'epidemia.⁵ Al fine di quantificare la capacità di propagazione di un'epidemia è, quindi, fondamentale monitorare le variazioni di R_0 che, in popolazioni sufficientemente numerose, può essere interpretato come il numero di nuovi infetti che si sostituiscono a un individuo che guarisce o muore anche in istanti temporali successivi a $t = 0$. Questa interpretazione consente di comprendere che la diffusione di un'epidemia può arrestarsi solo quando $R_0 < 1$. In tal caso, infatti, il numero di individui che escono dal compartimento degli infetti è sempre maggiore di quelli che vi entrano. Viceversa, $R_0 > 1$ implica che per ciascun individuo che guarisce o muore si ha più di un nuovo contagio.

Un modello in cui R_0 varia nel tempo tiene conto del fatto che il numero di nuovi infetti può cambiare non solo in ragione della diminuzione del bacino dei suscettibili, ma anche per motivi esterni alla dinamica stessa del modello di trasmissione, come quando, per effetto del distanziamento sociale, il numero medio di contatti di ciascun individuo si riduce e si ha una variazione nel tempo del parametro β e quindi di R_0 .

In questa analisi, per tenere conto della variazione di R_0 dovuta all'introduzione delle politiche di distanziamento interpersonale e/o ad altri fattori esterni (per esempio la variazione della temperatura), è stata assunta su questo parametro una funzione costante a tratti con primo salto il 16 marzo, ossia il lunedì successivo all'inizio del *lockdown*, e salti successivi a cadenza di 21 giorni fino al 18 maggio. Se r_1 , r_2 , r_3 , r_4 e r_5 rappresentano i valori medi del tasso netto di riproduzione nei cinque periodi considerati, l'andamento di $R_0(t)$ nel tempo qui assunto è:

$$R_0(t) = r_1 I \{t \in [t_{\min}, t_1]\} + \sum_{i=2}^4 r_i I \{t \in [t_{i-1}, t_i]\} + r_5 I \{t \in [t_4, t_{\max}]\}$$

dove t_{\min} corrisponde al 09.03.2020 e t_{\max} al 30.06.2020. Le assunzioni sottostanti il modello SIRD utilizzato sono numerose e devono essere tenute in considerazione nell'interpretare i risultati dell'analisi:

■ la popolazione è chiusa: non ci sono migrazioni, nascite e morti (a eccezione di quelle di individui affetti da Covid-19);

- ciascun individuo appartenente alla classe dei suscettibili ha la stessa probabilità di essere contagiato;
- una volta contratto il virus la probabilità di tornare a essere suscettibile è nulla, ovvero il tasso di perdita dell'immunità è pari a 0;
- ciascun individuo contagiato inizia a essere infettivo dall'istante successivo al contagio (in questo senso, i tempi T_d e T_r vanno intesi come comprensivi di un eventuale periodo di incubazione durante il quale il soggetto è infettivo);⁶
- a ogni istante temporale, la probabilità di contagiare un suscettibile è uguale per tutti gli infetti;
- i due soli possibili esiti della malattia sono la guarigione e la morte;
- la probabilità di guarire, così come quella di morire, è la stessa per tutti gli infetti;
- gli individui appartenenti alle classi dei guariti e dei morti non sono più infettivi.

IDENTIFICABILITÀ E PARAMETRIZZAZIONE DEL MODELLO

Nei modelli compartimentali si incorre spesso nel problema della non identificabilità dei parametri, che ha luogo quando combinazioni diverse dei parametri possono condurre a stesse dinamiche epidemiche. La non identificabilità può essere di tipo strutturale, non eliminabile, o di tipo pratico, quando è scarsa l'informazione su cui si basa l'inferenza, ma spesso le due componenti sono entrambe presenti.⁷ Nel modello SIRD specificato in questo studio, avendo a disposizione solo il dato relativo ai decessi, è semplice evincere che il tasso di mortalità e il tasso di guarigione non possano essere stimati, a meno di imporre un vincolo sulla numerosità del compartimento degli infetti e/o dei guariti. Infatti, diverse combinazioni di letalità e tempi alla guarigione e alla morte possono produrre gli stessi tassi di transizione e, quindi, gli stessi valori predetti per la serie dei decessi. Il parametro β è identificabile e la sua stima consente di identificare $R_0(t)$ dati α e δ o, nel caso in cui $T_d = T_r = T$, dato T . Tenendo conto di questo e del fatto che il principale obiettivo di questo studio è stimare la funzione $R_0(t)$, è stato fissato $T_d = T_r = 14$ giorni e $p = 1,14\%$, in accordo alla stima di letalità ottenuta per l'Italia dall'Istituto per gli studi di politica internazionale.⁸⁻¹⁰ Usando questi valori, si è ottenuto $\delta = 0,0008$ e $\alpha = 0,0706$. I parametri incogniti del modello SIRD da stimare tramite calibrazione sono rimasti, dunque, la data di inizio epidemia (data della prima infezione) e il vettore $r = (r_1, r_2, r_3, r_4, r_5)$.

CALIBRAZIONE

Diverse combinazioni dei parametri che regolano le variazioni della numerosità dei compartimenti nel corso del tempo corrispondono a diversi scenari evolutivi dell'epidemia. Calibrare significa cercare la combinazione di parametri ottimale che conduce a dinamiche epidemiche simili a quella realmente osservata. Nella presente analisi, è

stata effettuata una calibrazione basata sul confronto tra la numerosità del compartimento dei deceduti simulata dal SIRD e i decessi notificati dal 9 marzo al 30 giugno 2020. Tra tutte le possibili combinazioni dei parametri incogniti, è stata scelta quella che minimizzava gli scarti quadratici medi tra la cumulata dei decessi osservata e la cumulata dei decessi predetta dal modello. Tale minimizzazione è stata effettuata ricorrendo alla procedura di minimizzazione *hjn* implementata nel pacchetto *optimr* del software R.^{11,12} Come noto, in questo tipo di procedure è cruciale l'inizializzazione dell'algoritmo per evitare di incorrere in minimi locali.^{7,13} Pertanto, i punti iniziali sono stati scelti attraverso una minimizzazione preliminare della funzione obiettivo su una griglia multidimensionale poco fitta.

BOOTSTRAP PARAMETRICO

In questo studio, alle stime puntuali dei parametri sono stati associati gli intervalli di confidenza al 95% ottenuti con una procedura di *bootstrap* parametrico.¹⁴ Prima si sono ottenute 500 repliche della serie cumulata dei decessi campionando da distribuzioni binomiali negative definite sugli incrementi giornalieri della mortalità Covid-19 con media pari agli incrementi stimati dal modello SIRD. Quindi, ciascuna replica è stata considerata come serie osservata, rispetto alla quale effettuare la calibrazione del modello. In questo modo si sono ottenute 500 stime *bootstrap* per ciascun parametro incognito. Gli intervalli di confidenza sono stati costruiti in base ai percentili delle 500 stime *bootstrap*.

I parametri della distribuzione binomiale negativa sono stati scelti seguendo l'approccio di Grenfell e colleghi,¹⁵ che consente di tenere conto della sovradisersione, in particolare sugli incrementi giornalieri di bassa numerosità come quelli osservati all'inizio e alla fine della fase più emergenziale dell'epidemia.

SCENARI DI EVOLUZIONE A MEDIO E LUNGO TERMINE

A partire dal modello stimato sono state prodotte delle proiezioni a medio/lungo termine per la curva epidemica, ipotizzando semplici scenari definiti in termini di evoluzione di $R_0(t)$ nei mesi a venire. Le previsioni ottenute sono da leggersi in un'ottica qualitativa e servono per avere un'idea di massima di cosa potrebbe accadere se, a causa di fattori esterni, il numero di nuovi contagi generati da un individuo infetto dovesse aumentare rispetto al valore stimato al 30.06.2020. In particolare, sono state immaginate due tipologie di scenari:

- scenario 1: il valore medio di $R_0(t)$ raggiunto nel periodo dal 18 maggio al 30 giugno resta costante per un periodo sufficiente ad azzerare il numero degli infetti circolanti nella regione;
- scenario 2: il valore medio di $R_0(t)$ raggiunto nel periodo dal 18 maggio al 30 giugno resta costante fino al 30 settembre, per poi risalire oltre 1; $R_0(t) = 1,1$ (scenario 2a), 1,2 (scenario 2b), 1,3 (scenario 2c).

RISULTATI

Il modello SIRD ha individuato come data di inizio epidemia in Toscana il 21.02.2020. Come mostrato in figura 2 (grafico a sinistra), la combinazione di parametri del SIRD ottenuta dalla procedura di calibrazione ha generato una serie stimata dei decessi (in nero) fortemente coerente con la serie temporale dei decessi osservati in Regione nel periodo in studio (in rosso). In figura 2 (grafico a destra) sono riportati gli infetti stimati dal modello (in nero) e gli infetti diagnosticati (in rosso). Si osservi che gli infetti notificati sono molto inferiori agli infetti stimati dal modello che, per come costruito, tiene conto di tutti i contagi, inclusi quelli non rilevati. In particolare, si stima che alla fine di giugno il numero totale di infezioni sia stato 98.715, equivalente al 2,6% della popolazione regionale. Di queste infezioni, solo il 10% è stato notificato. In figura 3 è rappresentato l'andamento nel tempo della percentuale di infetti diagnosticati tra gli infetti circolanti stimati dal modello (in verde) e della percentuale di casi incidenti diagnosticati tra i nuovi casi stimati (in viola). Le due percentuali subiscono un forte incremento da fine aprile a metà maggio. In corrispondenza del picco epidemico, stimato attorno all'inizio di aprile, gli infetti circolanti diagnosticati risultano appena il 22% di quelli stimati.

Nel periodo *pre-lockdown*, il valore stimato di $R_0(t)$ è risultato pari a 7,78 (IC95% 7,55-7,89): fino al 15 marzo un infetto in media poteva trasmettere il virus a 7,8 persone. Nelle prime tre settimane di *lockdown* (16 marzo-5 aprile) $R_0(t)$ è calato bruscamente fino a 1,56 (IC95% 1,40-1,87) e nelle successive tre (6 aprile-26 aprile) è sceso a valori inferiori a 1 (0,8; IC95% 0,57-0,97), fino ad arrivare a 0,14 (IC95% 0-0,43) dal 27 aprile al 17 maggio.

Infine si è osservata una risalita dell'indice, stimato essere pari a 0,66 (IC95% 0,32-0,88) dal 18 maggio a fine giugno. L'andamento di $R_0(t)$ è riportato in figura 4.

Cambiando il valore della letalità, si ottengono stime della funzione sovrapponibili (risultati non riportati). Il parametro ha, invece, un'ovvia influenza sulla quantificazione del numero di infetti e quindi sulla percentuale di infezioni notificate sul totale di quelle stimate da modello. In particolare, all'aumentare/diminuire della letalità tale percentuale aumenta/diminuisce.

Per quanto riguarda le previsioni a medio-lungo termine (figura 5), lo scenario 1 assume un valore di $R_0(t)$ costante e pari a 0,66 fino alla data di completo esaurimento del bacino degli infetti, stimata al 29.04.2021. In questo scenario ottimistico, una volta azzerati gli infetti l'epidemia si può considerare esaurita, a meno di un ritorno dell'infezione dall'esterno, di cui il modello SIRD, che assume una popolazione chiusa, non tiene conto. Nei grafici in alto in figura 5 viene mostrato l'andamento delle curve epidemiche degli infetti circolanti (a sinistra) e dei decessi (a destra) sotto lo scenario 1. Il modello SIRD stima che il picco di infetti circolanti in regione si sia raggiunto attorno ai primi di aprile (poco meno di 30.000 infetti circolanti) e che, restando l' $R_0(t)$ costante e pari a 0,66, il numero totale dei decessi sarà pari a circa 1.170, valore raggiunto a metà febbraio 2021. Sotto lo scenario 1, il numero totale di contagiati durante tutto il corso dell'epidemia sarebbe pari a circa 100.000 individui.

I grafici in basso in figura 5 mostrano gli infetti circolanti (a sinistra) e i decessi (a destra) assumendo che dal 1 ottobre $R_0(t)$ sia pari a 1,1, 1,2 e 1,3 (scenari 2a, 2b e 2c). La prima cosa che si nota è come valori di R_0 così simili pro-

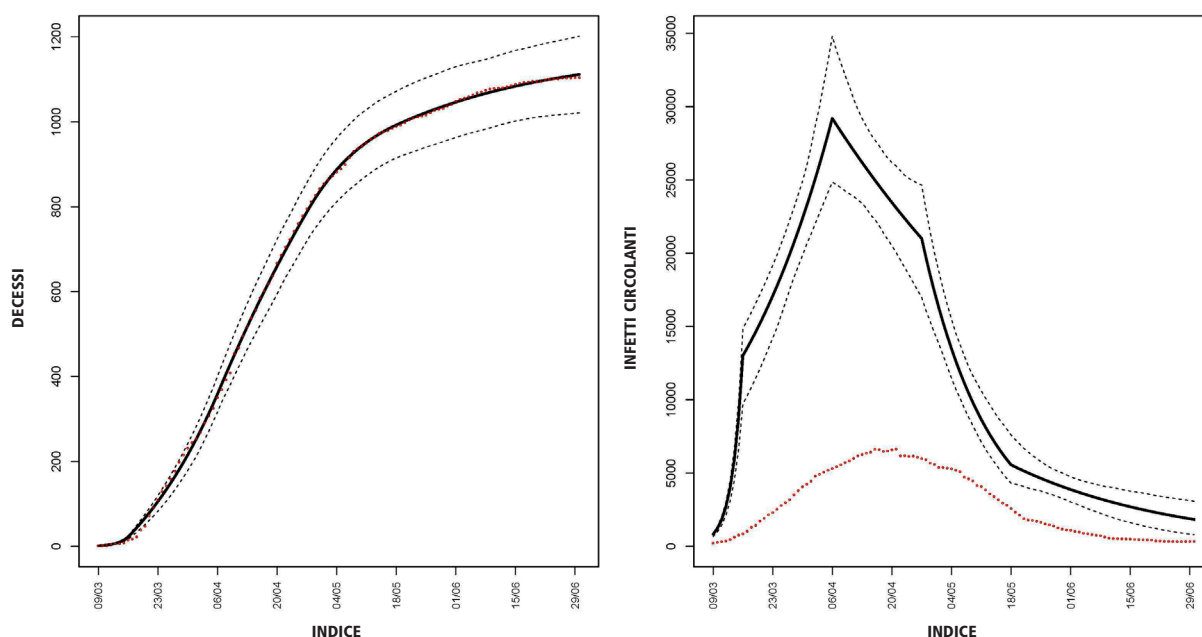


Figura 2. Serie stimate e bande di confidenza al 95% ottenute dal modello SIRD (linee nere) e serie osservate (punti rossi) dei decessi cumulati (sinistra) e degli infetti circolanti (destra).

Figure 2. Estimated time series and 95% confidence bands obtained from the SIRD model (black) and observed time series (red) of cumulative deaths (left panel) and circulating infected (right panel).

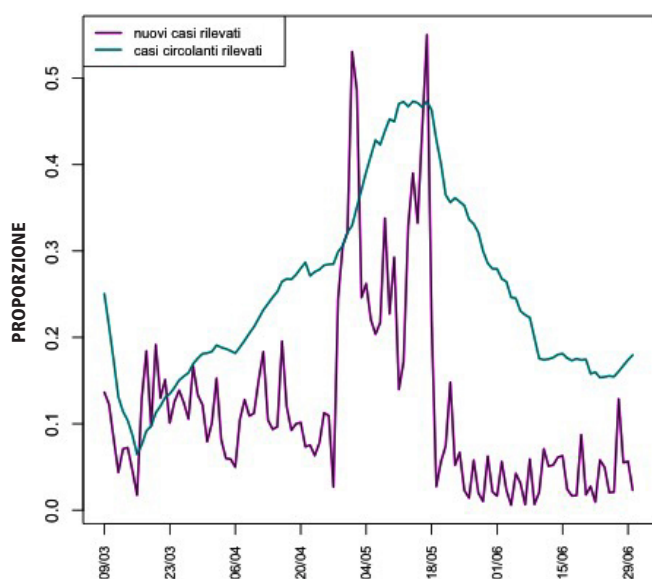


Figura 3. Percentuale di nuove infezioni notificate sul totale delle nuove infezioni stimate dal modello e percentuale di infezioni notificate circolanti in regione sul totale delle infezioni circolanti stimate da modello.

Figure 3. Percentage of new notified infections over the total number of new infections estimated by the model, and percentage of notified infections circulating in the Region over the total number of infections circulating in the Region estimated by the model.

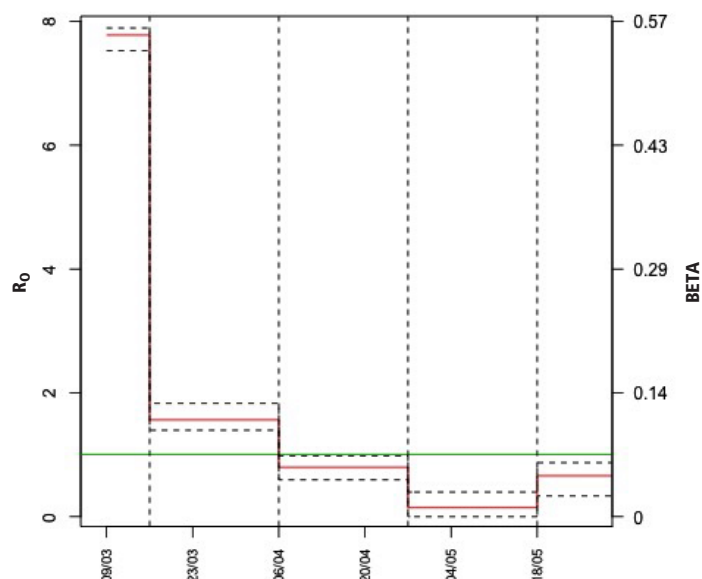


Figura 4. Andamento dell'indice $R_0(t)$ (asse delle ordinate sul lato sinistro) e del parametro $\beta(t)$ (asse delle ordinate sul lato destro) nel tempo. Stima puntuale (linea rossa) e intervalli di confidenza al 95% (linee nere tratteggiate). In verde il valore soglia $R_0(t)=1$

Figure 4. Estimate of $R_0(t)$ (left axis) and $\beta(t)$ (right axis). Point estimate (red line), 95% confidence interval (dashed black lines), threshold $R_0(t)=1$ (green line).

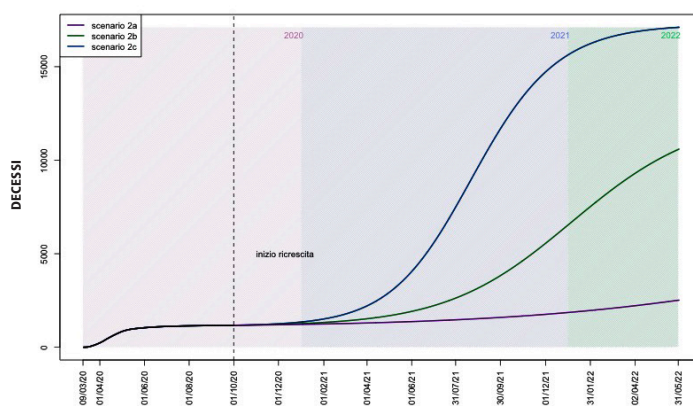
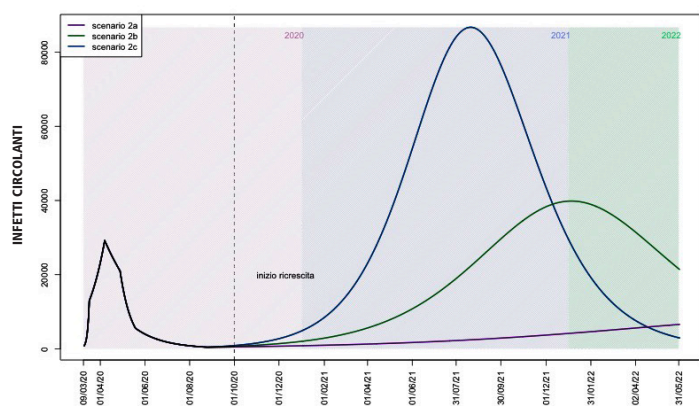
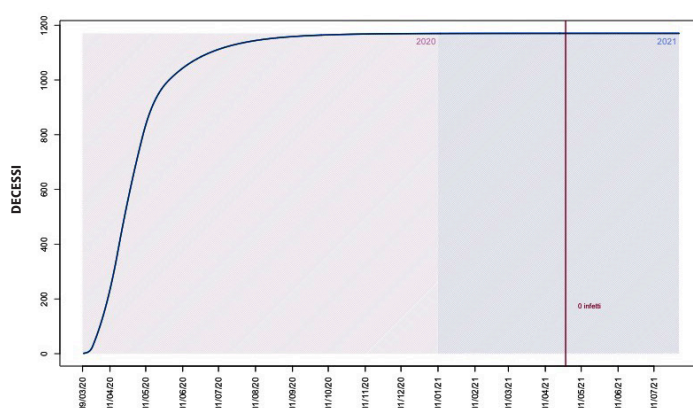
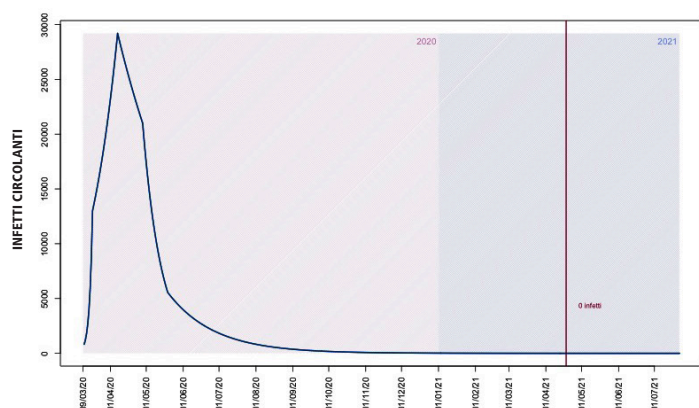


Figura 5. Evoluzione delle curve epidemiche in scenari a lungo termine: infetti circolanti (a sinistra) e decessi (a destra). Scenario 1 (in alto): $R_0(t) = 0.66$ dal 30 giugno fino alla data di completo esaurimento del bacino degli infetti. Scenario 2 (in basso): $R_0(t) = 1.1$ (scenario 2a), $R_0(t) = 1.2$ (scenario 2b), $R_0(t) = 1.3$ (scenario 2c) dal 1 ottobre 2020.

Figure 5. Epidemic curves under long-term scenarios: circulating infected (left panel) and cumulative deaths (right panel). Scenario 1 (upper panel): $R_0(t) = 0.66$ from June 30th until the complete exhaustion of the infected. Scenario 2 (lower panel): $R_0(t) = 1.1$ (scenario 2a), $R_0(t) = 1.2$ (scenario 2b), $R_0(t) = 1.3$ (scenario 2c) from October 1st, 2020.

ducano dinamiche epidemiche completamente diverse. Se sotto lo scenario 2a la crescita del numero degli infetti circolanti è molto lenta, negli altri due casi si osservano picchi di ritorno dell'epidemia. In particolare, sotto lo scenario 2c l'onda del contagio potrebbe raggiungere livelli di nuovo preoccupanti entro la prossima primavera. Ovviamente, qualora dal 1 ottobre $R_0(t)$ assumesse valori ancora più elevati di quelli previsti nello scenario 2, e/o iniziasse a crescere prima di questa data (quindi con un numero di infetti circolanti maggiore), l'onda del contagio sarebbe anticipata e di maggiore entità.

DISCUSSIONE

Il modello SIRD utilizzato ha consentito di stimare l'evoluzione di $R_0(t)$ in Toscana, dall'inizio dell'epidemia fino alla fine di giugno 2020. L'analisi ha evidenziato una sostanziale decrescita del tasso netto di riproduzione fino al 18 maggio e una sua successiva risalita che comunque non ha portato al superamento della soglia critica di 1. Anche se non è stato stimato direttamente l'impatto del *lockdown* su $R_0(t)$, questo risultato sembra suggerire che vi è stato un forte effetto delle politiche di contenimento sulla diffusione del virus in Toscana e di una minima ripresa del contagio potenzialmente attribuibile alla riapertura, avvenuta il 04.05.2020.

La scelta delle finestre temporali per la funzione costante a tratti utilizzata per la stima di $R_0(t)$ è stata effettuata a priori. Distanziare gli scalini di 21 giorni ha consentito di ottenere stime stabili cogliendo al contempo l'andamento complessivo del tasso di riproduzione. Come alternativa si sarebbe potuta definire su $R_0(t)$ una *spline* di regressione, ma la forte sensibilità dell'algoritmo di ottimizzazione ai valori iniziali dei parametri ha fatto optare per una soluzione più semplice.

Il modello SIRD specificato in questo studio consente di ottenere una stima del numero totale di infezioni, incluse quelle non diagnosticate. Assumendo una letalità pari all'1,14%, gli infetti notificati e riportati nel data set della Protezione civile rappresentano una piccola parte – meno del 25% – di quelli totali stimati dal modello.¹⁰ Tuttavia, questa percentuale tende a crescere da fine aprile a metà maggio. Se, da una parte, questo andamento può suggerire che la capacità di diagnosticare le infezioni è aumentata in vista della riapertura e nel periodo immediatamente successivo, dall'altra è importante notare che il picco nella percentuale di casi diagnosticati e circolanti è esattamente in corrispondenza della finestra temporale dove $R_0(t)$ è stimato essere molto vicino a 0. Pertanto, i valori percentuali più elevati potrebbero essere dovuti a un denominatore molto basso che rende la stima stessa della proporzione piuttosto instabile.

Riguardo alla scelta del valore della letalità, questa potrà essere affinata alla luce di eventuali risultati da studi campionari di siero-prevalenza, preferibilmente condotti a livello locale. Il parametro potrà essere in questo caso stimato attraverso una calibrazione tra il numero totale di infezioni

predetto dal modello SIRD e quello ottenuto dall'indagine campionaria condotta sulla popolazione.

Oltre a essere strumenti utili per la stima e il monitoraggio delle dinamiche epidemiche, i modelli compartimentali consentono di fare simulazioni a medio-lungo termine, grazie alla loro forte componente meccanicistica. In questo articolo, sono state proiettate le dinamiche epidemiche nei prossimi mesi, ipotizzando differenti scenari relativi all'andamento di $R_0(t)$. Gli altri parametri del modello, in particolare il tasso di guarigione e il tasso di letalità per Covid-19, sono stati, invece, lasciati costanti, a indicare che si sta assumendo che la letalità dell'infezione e i tempi di guarigione o morte dei malati non siano soggetti a cambiamento.

Le proiezioni ottenute devono essere interpretate in modo qualitativo e con la dovuta cautela, anche tenendo presente che, più lungo è l'orizzonte temporale, maggiori sono i fattori di incertezza che possono entrare in gioco. Tuttavia, esse mostrano inequivocabilmente come il pericolo non sia scongiurato ed evidenziano la necessità di essere prudenti, dal momento che anche valori di $R_0(t)$ ampiamente al di sotto di quello stimato nelle fasi iniziali dell'epidemia possono provocare nuove ondate di contagio entro la prossima primavera.

Questi scenari devono essere pensati come ascrivibili a variazioni in uno o più fattori che abbiano un impatto sul numero medio di contatti di ciascun individuo infetto e sulla probabilità che un contatto possa generare una nuova infezione. Il distanziamento fisico, la diffusione e l'utilizzo di dispositivi per la protezione individuale, la capacità del virus di contagiare – che potrebbe essere legata a fattori ambientali e/o seguire un andamento stagionale – e la capacità dei sistemi di sorveglianza di tracciare e isolare precocemente i soggetti infetti, sono determinanti del tasso di trasmissione e quindi delle variazioni di $R_0(t)$. Laddove fosse possibile una formalizzazione affidabile della relazione tra questi fattori e $R_0(t)$, anziché ipotizzare semplici variazioni del tasso di riproduzione come è stato fatto in questo articolo, si potrebbero fare proiezioni delle dinamiche epidemiche sotto scenari realistici relativi al contesto sociale, ambientale e sanitario in cui si sviluppa e diffonde il contagio. E, ancora, avendo a disposizione stime causali dell'impatto di politiche di contenimento dell'epidemia o allentamento delle misure restrittive sul comportamento dei cittadini, il tipo di modellazione proposto potrebbe essere di ausilio per pianificare azioni future. Tuttavia, in mancanza di dati affidabili da cui ricavare le suddette informazioni, ogni scenario che scenda in dettagli non supportati da evidenze scientifiche ha, a parere di chi scrive, carattere altamente speculativo. Nell'ottica di una possibile ripresa del contagio da SARS-CoV-2 nei prossimi mesi, ma anche al fine di non essere colti impreparati da emergenze simili che dovessero presentarsi nel futuro, è importante accrescere le conoscenze relativamente alla risposta della popolazione alle misure di distanziamento fisico, all'effi-

cacia dell'utilizzo di dispositivi di protezione individuale e delle strategie di tracciamento precoce degli infetti, attraverso studi di tipo osservazione o sperimentale.

L'approccio di questo studio, per molti aspetti simile a quello proposto da altri,¹⁶⁻¹⁸ ha delle caratteristiche peculiari. Rispetto ai modelli che si concentrano su dati a livello nazionale o di macroarea, l'utilizzo di dati regionali consente di ridurre l'eterogeneità dei parametri che regolano l'epidemia (tra questi anche la data di inizio del contagio), ottenendo stime e proiezioni più aderenti al contesto locale, quindi potenzialmente più affidabili. La calibrazione è effettuata sui decessi e non sugli infetti notificati, che rappresentano solo una parte dei contagi effettivi. Alle stime di questo studio è stata associata una valutazione della variabilità attraverso il calcolo di intervalli di confidenza tramite una procedura di *bootstrap* parametrico.

L'approccio adottato presenta altresì alcune limitazioni. Prima di tutto, il modello SIRD si basa su ipotesi forti, che riducono 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 $R_0(t)$ 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). Il modello non tiene altresì conto di una possibile variazione temporale dei parametri che sono stati tenuti fissi per motivi di identificabilità (letalità e tempo di morte e guarigione) e non incorpora al suo interno l'incertezza legata alla scelta dei valori a essi assegnati, e la sua ripercussione sulla stima del tasso netto di riproduzione.

Questa ultima considerazione si applica in particolare al tempo di risoluzione dell'infezione T , dal momento che la stima di $R_0(t)$ è risultata robusta alla scelta di p .

Una debolezza dell'approccio utilizzato in questo lavoro risiede anche nell'utilizzo di un'unica serie di dati per la calibrazione, quella dei decessi. Questa scelta ha consentito di evitare i problemi insiti nei dati di notifica di infezione e di guarigione, come quello delle registrazioni differite e dei ritardi nella notifica, che dovrebbero essere meno accentuati nel dato di mortalità, ma soprattutto ha consentito di usare un modello SIRD semplice, sfruttando la corrispondenza tra compartimento dei deceduti e serie osservata. Volendo usare in fase di calibrazione le serie di infetti e guariti, sarebbe venuta meno la corrispondenza tra i compartimenti I e R e le corrispondenti serie osservate, dal momento che i casi notificati sono solo una porzione del totale, mentre il modello SIRD ricostruisce infetti e guariti effettivi. La soluzione per incorporare informazione aggiuntiva sta nella formulazione di modelli più complessi, con un maggior numero di compartimenti e parametri di transizione, o in aggiustamenti ad hoc in fase di calibrazione.

Conflitti di interesse dichiarati: nessuno.

Contributo degli autori: definizione del modello e dei metodi, analisi statistiche e scrittura dei codici in R: MB, GC, CV.

Supervisione delle analisi statistiche e coordinamento: MB.

Discussione dei risultati e revisione della letteratura: MB, GC, CV, FM, AG.

Scrittura della prima bozza dell'articolo: MB, GC, CV, FM.

Tutti gli autori hanno letto e approvato la versione finale dell'articolo.

Data di sottomissione: 14.07.2020

Data di accettazione: 06.08.2020

BIBLIOGRAFIA

- Murray JD. Mathematical Biology. I: An Introduction. Berlin, Springer, 2002.
- Lin F, Muthuraman K, Lawley M. An optimal control theory approach to non-pharmaceutical interventions. BMC Infect Dis 2010;10:32.
- Maltagliati M. Covid 19: proviamo a prevedere cosa succederà. Neodemos: popolazione società e politica 2020. Disponibile all'indirizzo: <https://www.neodemos.info/articoli/covid-19-proviamo-a-prevedere-cosa-succederà/>
- Legrand J, Grais RF, Boelle PY, Valleron AJ, Flahault A. Understanding the dynamics of Ebola epidemics. Epidemiol Infect 2007;135(4):610-21.
- Flaxman S, Mishra S, Gandy A et al. Estimating the effects of non-pharmaceutical interventions on Covid-19 in Europe. Nature 2020;584(7820):257-61.
- Lauer SA, Grantz KH, Bi Q et al. The incubation period of coronavirus disease 2019 (Covid-19) From publicly reported confirmed cases: estimation and application. Ann Intern Med 2020;172(9):577-82.
- Roosa K, Chowell G. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. Theor Biol Med Model 2019;16(1):1.
- Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020;382(18):1708-20.
- Istituto superiore di sanità, Gruppo della sorveglianza Covid-19. Caratteristiche dei pazienti deceduti positivi all'infezione da SARS-CoV-2 in Italia. Roma, ISS, 2020. Disponibile all'indirizzo: <https://www.epicentro.iss.it/coronavirus/sars-cov-2-decessi-italia>
- Villa M. Coronavirus: la letalità in Italia, tra apparenza e realtà. Milano, ISPI, 2020. Disponibile all'indirizzo: <https://www.ispi.org/coronavirus-la-letalita-italia-tra-apparenza-e-realta-25563>
- R core team. R: A language and environment for statistical computing. Vienna, R Foundation for Statistical Computing, 2019. Disponibile all'indirizzo: <https://www.R-project.org/>
- Nash JC. Compact numerical methods for computers: linear algebra and function minimisation. Bristol, UK, Institute of Physics Publications, 1990.
- Zucchini W, MacDonald IL, Langrock R. Hidden Markov models for time series: an introduction using R. Second edition. Boca Raton, FL, USA, Chapman and Hall – CRC press, 2016.
- Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. Infect Dis Model. 2017;2(3):379-98.
- Grenfell BT, Bjørnstad ON, Finkenstädt BF. Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model. Ecological Monographs 2002;72(2):185-202.
- Fanelli D, Piazza F. Analysis and forecast of Covid-19 spreading in China, Italy and France. Chaos Solitons & Fractals 2020;134:109761.
- Caccavo D. Chinese and Italian Covid-19 outbreaks can be correctly described by a modified SIRD model. medRxiv 2020;2020.03.19.20039388. doi: 10.1101/2020.03.19.20039388
- Loli Piccolomini E, Zama F. Monitoring Italian Covid-19 spread by an adaptive SEIRD model. medRxiv 2020;2020.04.03.20049734; doi: 10.1101/2020.04.03.20049734