

Stime di parametri di modelli epidemiologici SIR semplificati utilizzando Excel 2016 relativi ai dati di diffusione della COVID-19 in Italia

Introduzione

Il massiccio flusso mediatico di informazioni relativi alla pandemia da Covid-19 rappresenta un'occasione per studenti ed insegnanti per affrontare temi raramente dibattuti in condizioni ordinarie e avvicinare gli stessi a concetti di statistica non ordinari.

Obiettivi

L'obiettivo dello studio è quello di aiutare gli studenti a pensare alla diffusione della Covid-19 in termini quantitativi analizzando i dati italiani del contagio e ricavando, attraverso gli strumenti di Solver e Calcolo di Excel 2016, informazioni su parametri epidemiologici desunti da modelli SIR semplificati i cui andamenti nel tempo riflettono l'efficacia delle misure di contrasto intraprese (lockdown, contact tracing, distanziamento sociale, mascherina).

Metodi e procedure

I dati italiani* relativi alla prima ondata vengono confrontati con un modello SIR semplificato¹. A tal fine, il pieno sviluppo di una curva pandemica viene suddivisa in tre fasi:

- la prima fase è rappresentata da un modello di crescita dei contagi esponenziale (CE) $R_i = \frac{\Delta I}{\Delta t} = \beta_i N_0 e^{\beta_i t}$ (1) dove:
 R_i = casi positivi giornalieri;
 Δt = 1 giorno;
 $I(t_i)$ = numero infetti al giorno t_i ;
 N_0 = numero infetti al tempo iniziale;
 β_i = tasso di contagio.
- la seconda fase riflette i primi effetti delle misure di contenimento della diffusione. Tale fase è esclusa dallo studio quantitativo.
- la terza fase è rappresentata da un modello di decremento dei contagi esponenziale (DE) $R_i = \frac{\Delta I}{\Delta t} = \beta_d N_0 e^{-\beta_d t}$ (2).

Si analizzano anche i dati di letalità apparente giornaliera relativi alla prima e seconda ondata col seguente modello:

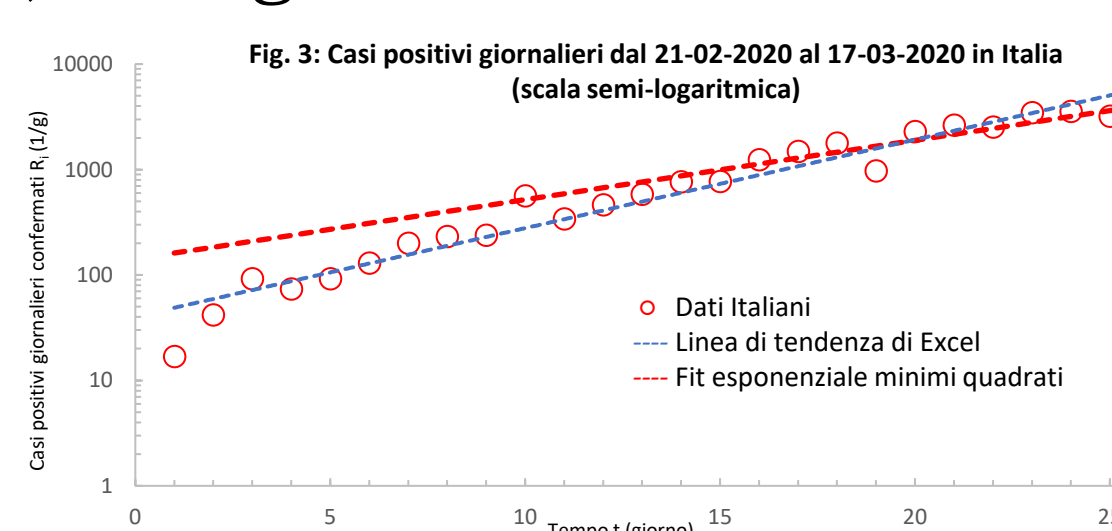
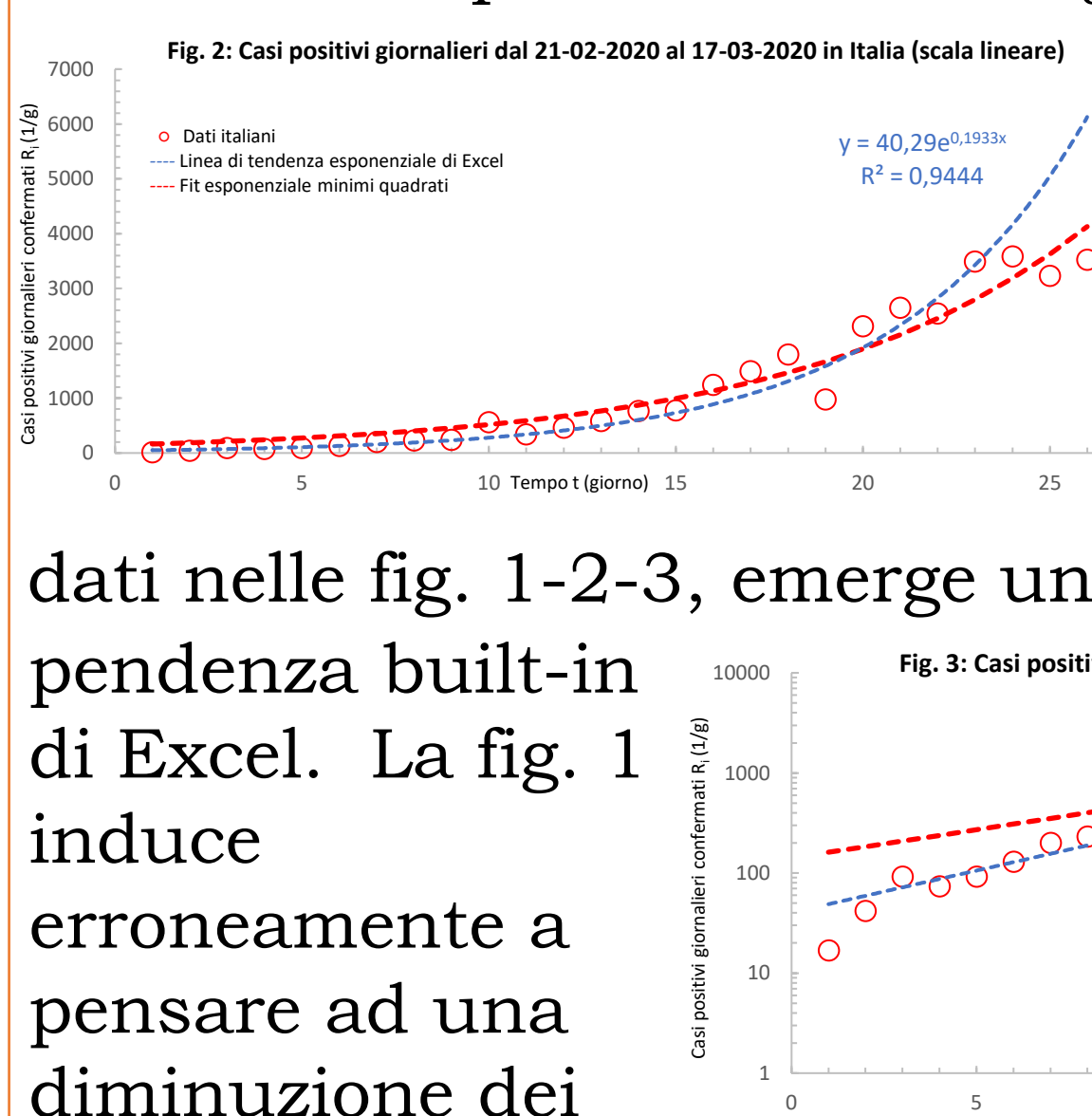
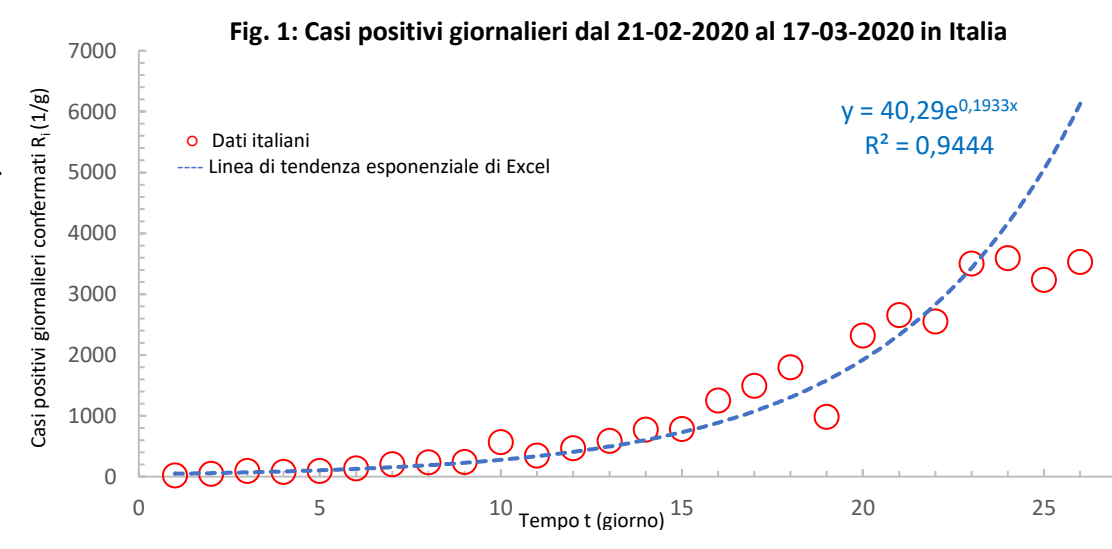
$$R_m(t) = m_r R_i(t - t_0) \quad (3) \text{ dove:}$$

m_r = tasso di letalità apparente;

t_0 = tempo medio tra la positività e il decesso.

Tutti i dati vengono analizzati con la procedura dei minimi quadrati con lo strumento Solver di Excel con parametri liberi $N_0, \beta_i, \beta_d, m_r$, con Q

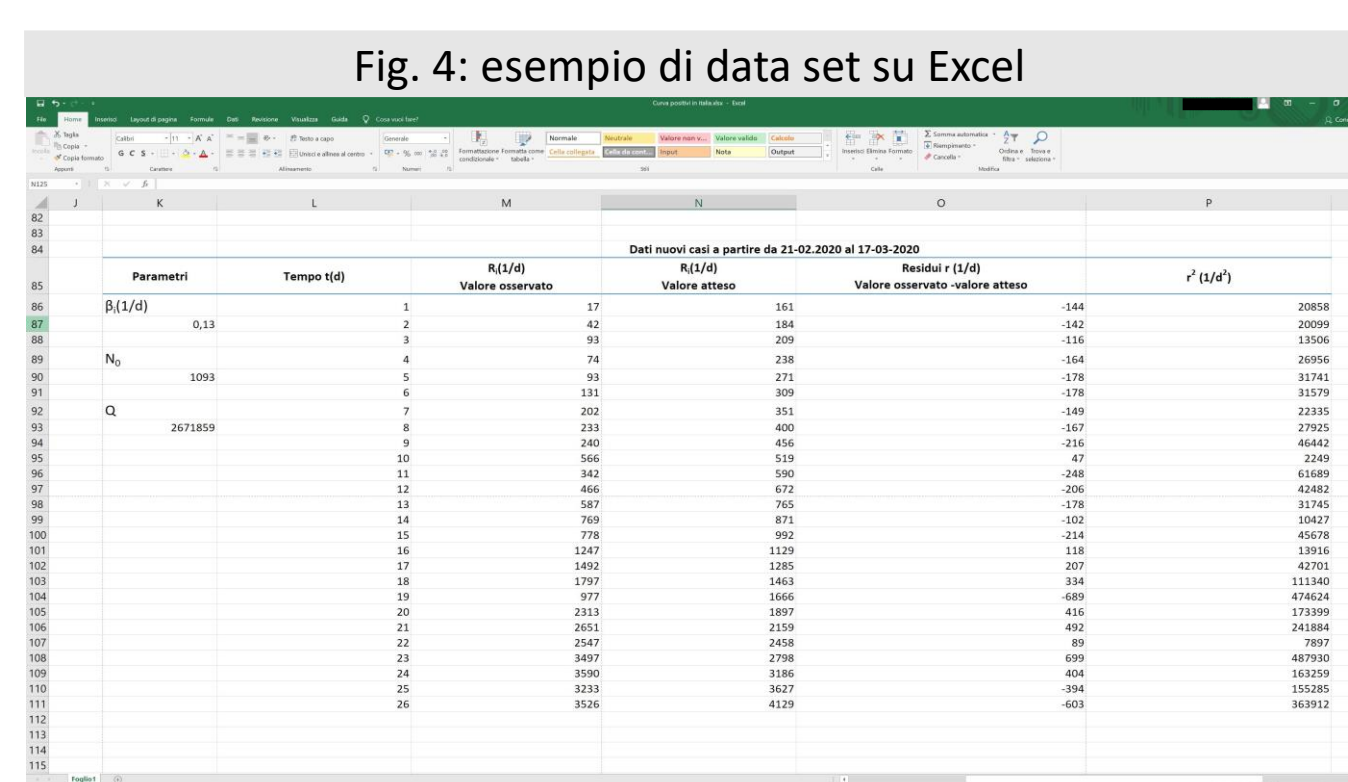
somma dei residui al quadrato come parametro obiettivo. Da una valutazione preliminare dei dati nelle fig. 1-2-3, emerge un bias nella linea di pendenza built-in di Excel. La fig. 1 induce erroneamente a pensare ad una diminuzione dei contagi negli ultimi tre giorni rispetto alle previsioni del modello. Le fig. 2 e 3 dimostrano la presenza del bias del fit esponenziale che, in Excel, è una regressione lineare sul logaritmo dei dati,



per come si evince anche dalla fig. 3 in scala semi-logaritmica, che produce una maggiore "propensione" a rappresentare meglio i dati nella prima fase di contagio al contrario del best-fit realizzato con Solver, in cui i residui sono omogeneamente distribuiti rispetto alla curva esponenziale, prestandosi meglio ad una analisi a tempi più grandi quando le misure di contenimento iniziano a riflettersi sull'andamento del modello teorico (appiattimento della curva dei contagi).

Risultati

La fig. 4 mostra un esempio di tabella dati analizzata con Excel. La fig. 5 rappresenta la curva pandemica italiana relativa alla prima ondata. Vengono visualizzati i casi positivi giornalieri, evidenziando le tre fasi



della curva pandemica con tre colori diversi. I dati iniziano dal 21-02-2020 e viene riportata la linea di tendenza di crescita esponenziale relativa ai primi 26 giorni, la sua equazione e l' R^2 . Lo

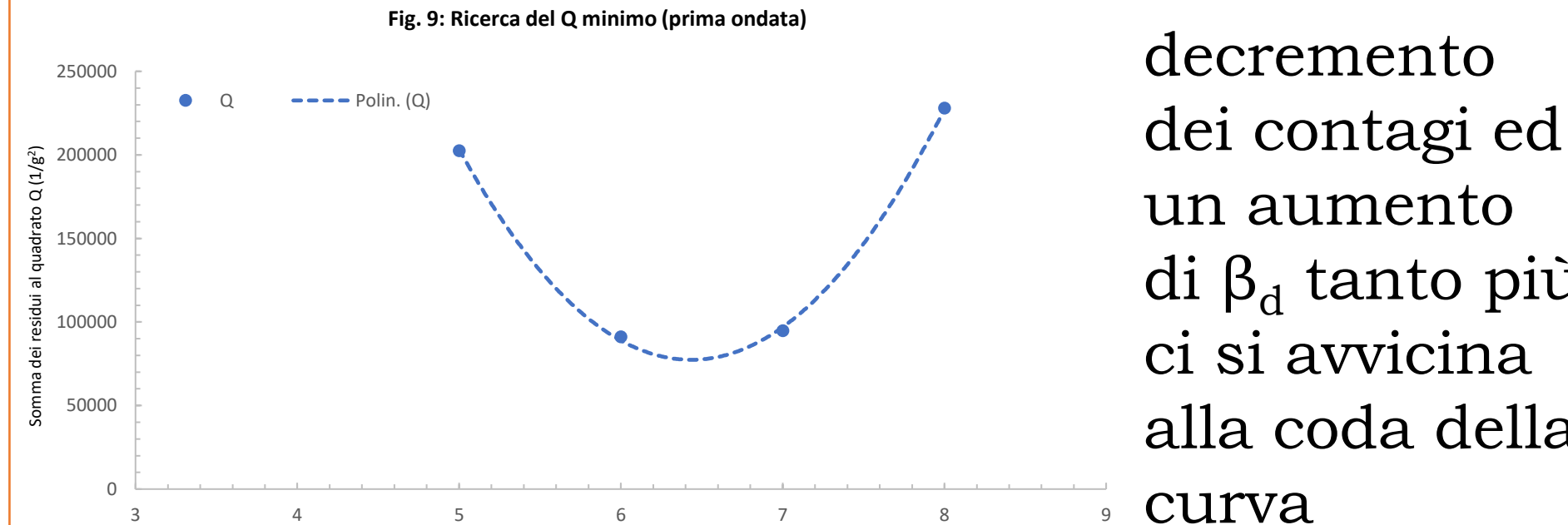
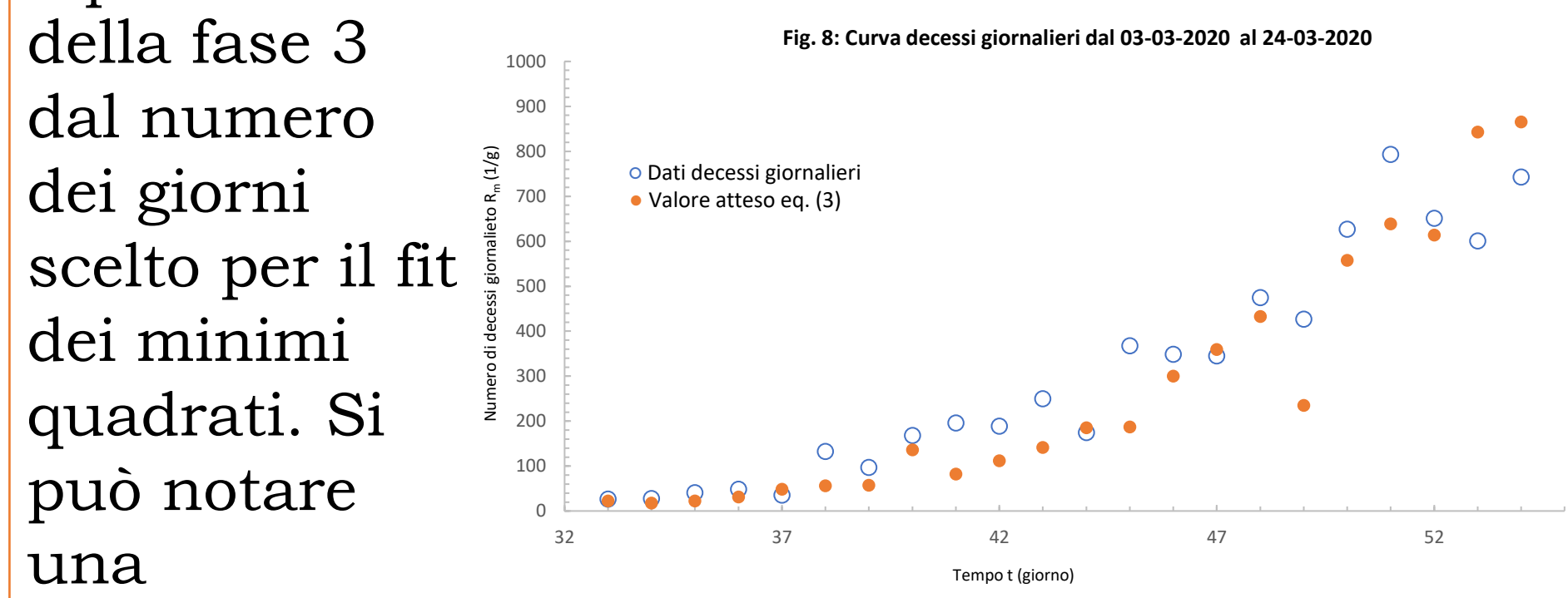
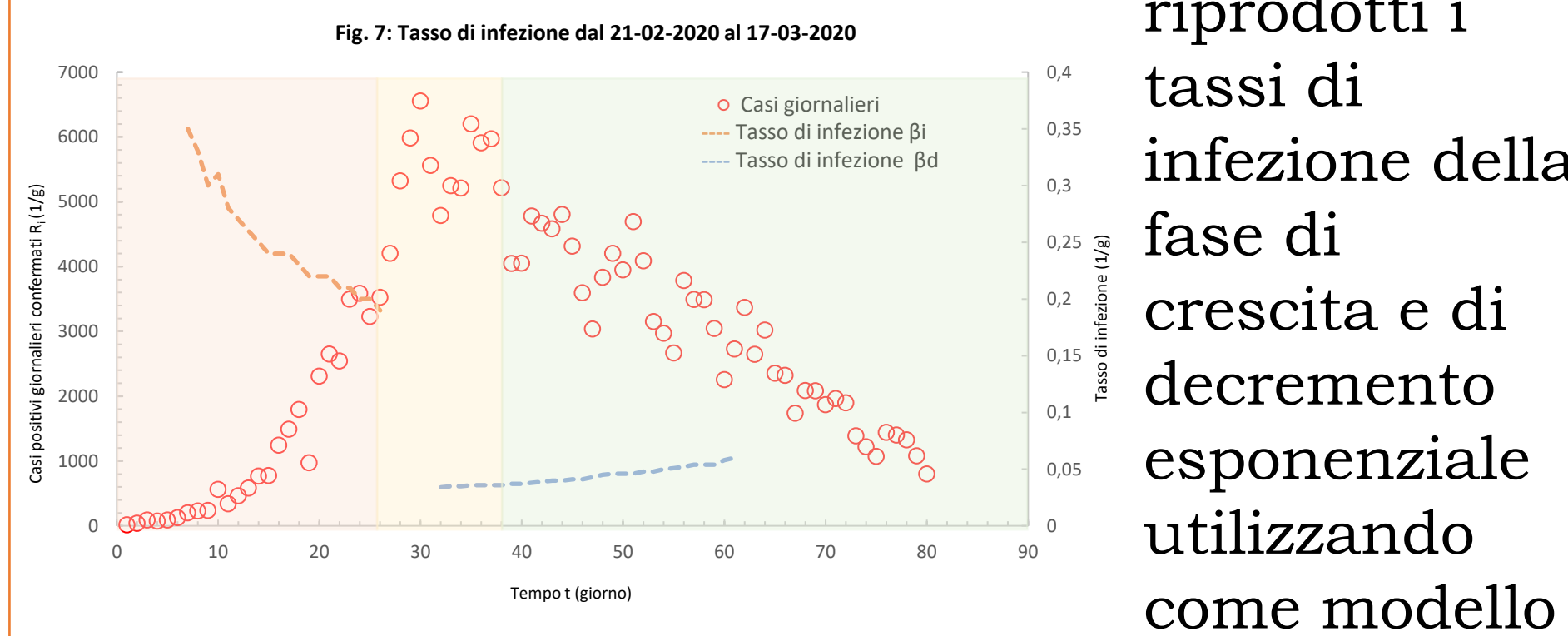
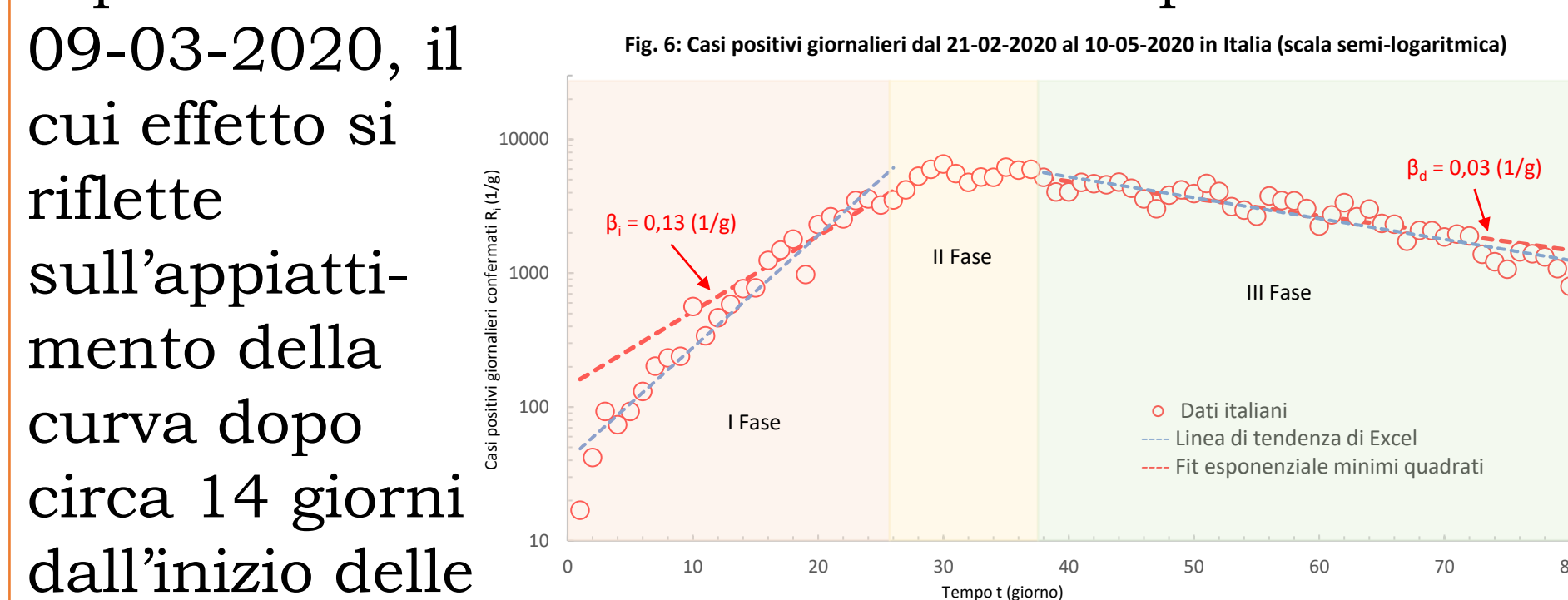
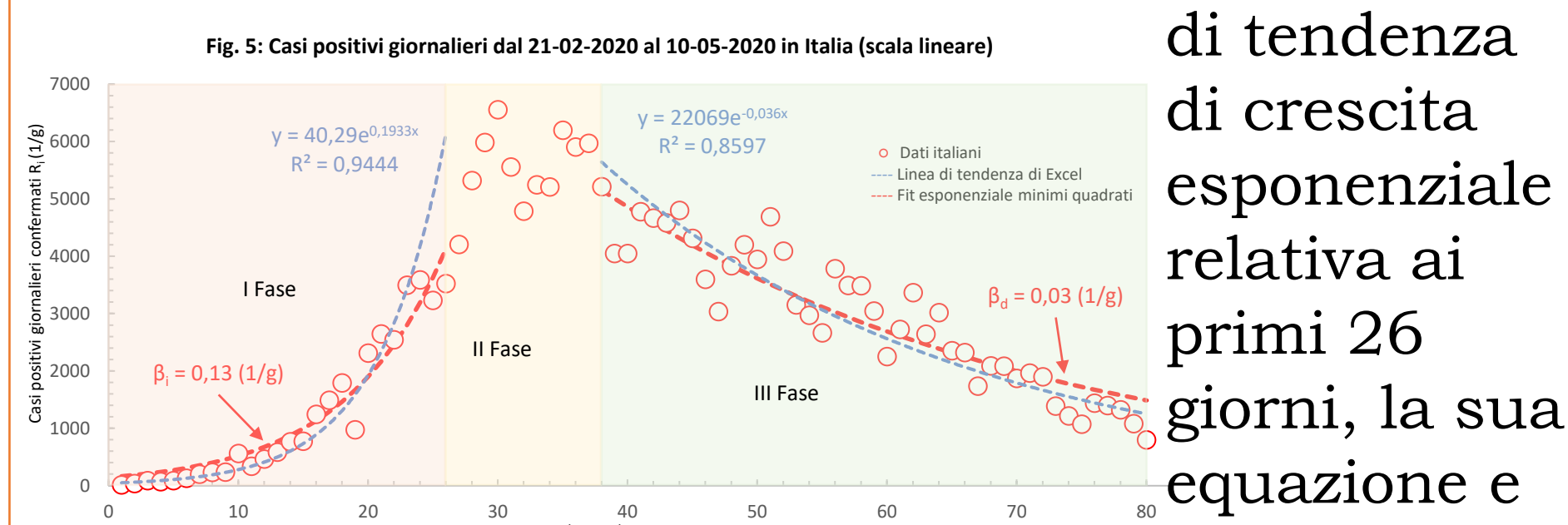
stesso viene fatto per la fase di decremento esponenziale successiva al lockdown proclamato il 09-03-2020, il cui effetto si riflette sull'appiattimento della curva dopo circa 14 giorni dall'inizio delle

restrizioni. La fig. 6 mostra gli stessi dati in scala semi-logaritmica. In entrambi i grafici vengono riprodotti i tassi di infezione della fase di crescita e di decremento esponenziale utilizzando come modello

teorico l'eq.ni (1) e (2). La fig. 7 mostra la dipendenza dei tassi di infezione della fase 1 e della fase 3 dal numero dei giorni scelti per il fit dei minimi quadrati. Si può notare una

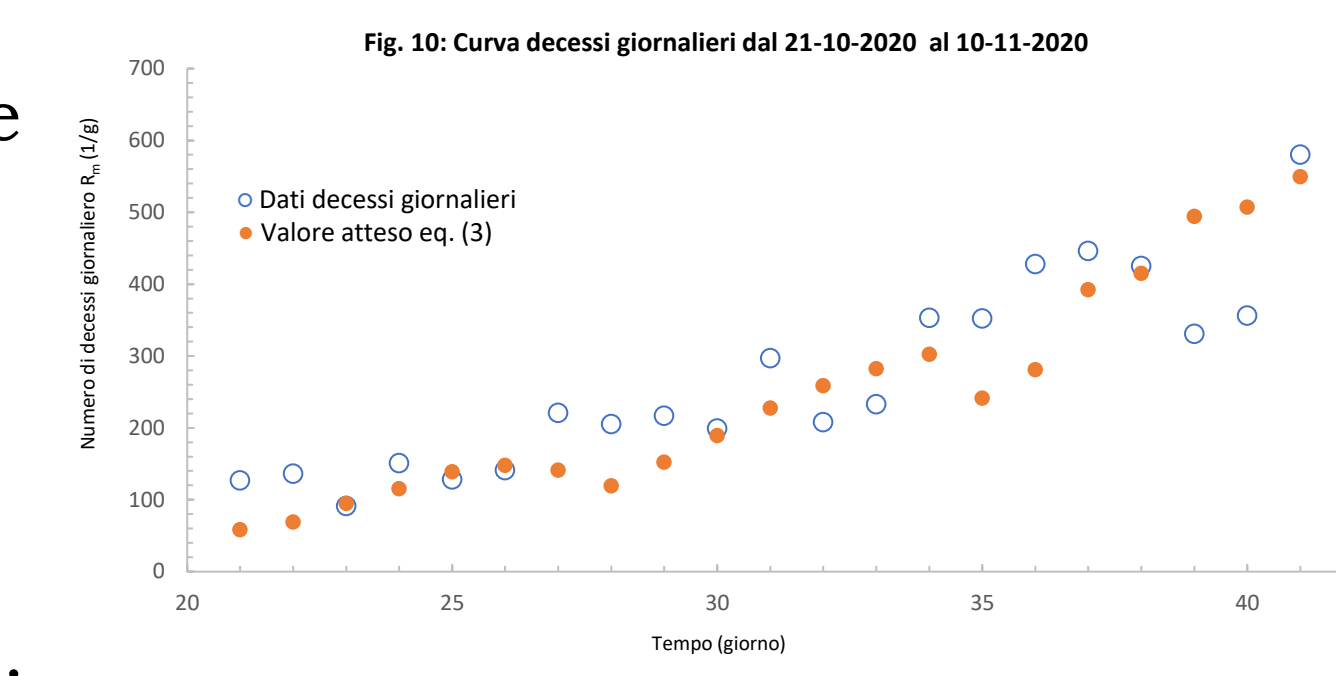
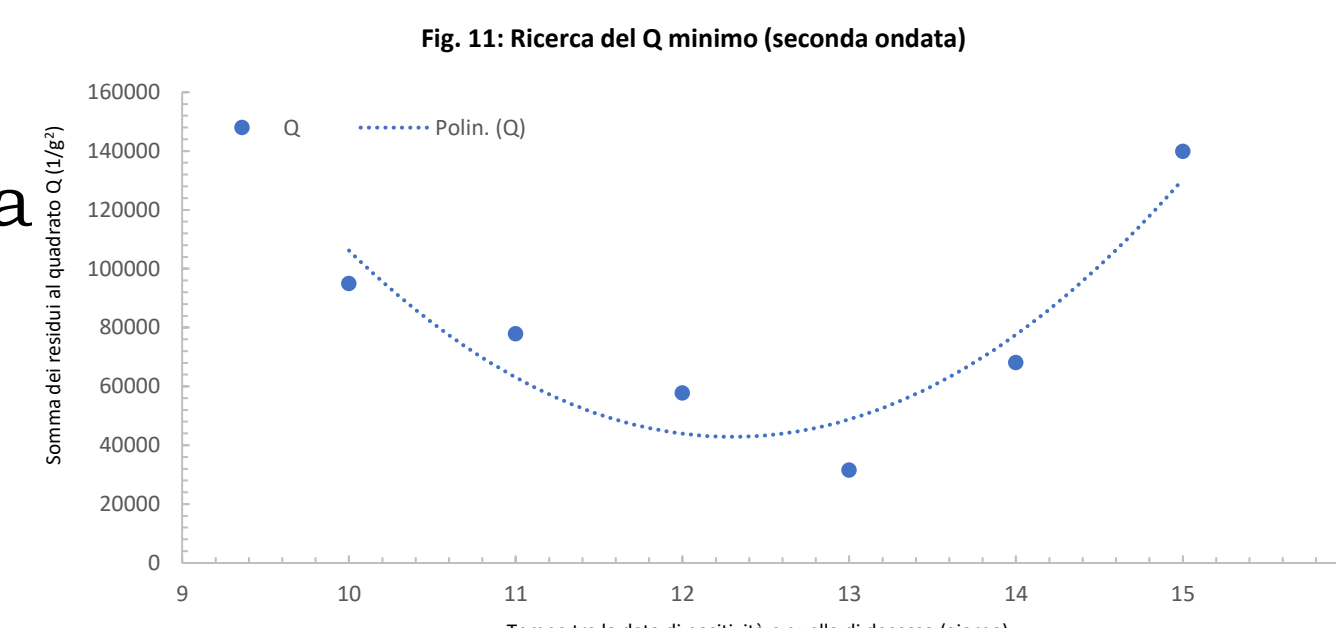
diminuzione del β_i all'avvicinarsi ai tempi della seconda fase in cui avviene la transizione crescita-decremento dei contagi ed un aumento di β_d tanto più ci si avvicina alla coda della curva

pandemica. La fig. 8 mostra un confronto tra i dati dei decessi giornalieri relativi alla fase di crescita esponenziale durante la prima ondata ed il valore teorico atteso dato dall'eq. (3) in cui m_r viene stimato attraverso una procedura di minimizzazione dei residui al quadrato.



La fig. 9 ci consente di stabilire il tempo medio t_0 tra il giorno di positività e quello di decesso osservando il valore in corrispondenza del quale si ottiene il Q minore. Sul grafico viene rappresentata

anche la linea di tendenza quadratica restituita da Excel. Per la prima ondata, la tab. 1 restituisce un valore intorno al 6° giorno. Le fig. 10-11 mostrano la stessa analisi relativamente alla seconda ondata. La tab. 2 mostra un tempo medio t_0 intorno al tredicesimo giorno.



t_m (g)	m_r	Q (1/g ²)
5	0,16	202580
6	0,18	91074
7	0,2	94791
8	0,22	228003

t_m (g)	m_r	Q (1/g ²)
10	0,016	94881
11	0,018	77827
12	0,020	57722
13	0,021	31502
14	0,023	68050
15	0,024	139906

Conclusioni

Attraverso modelli semplici ma potenti per descrivere la diffusione della COVID-19, si possono analizzare gli andamenti e gli effetti delle misure di contenimento attraverso parametri appropriati incorporati nel modello per la valutazione e la previsione quantitativa.

Le stime dei tassi di infezione delle fasi di crescita e decremento esponenziali consentono di monitorare lo sviluppo dell'epidemia, riflettendo fattori pandemici intrinseci ed esterni, sono compatibili con altri lavori di letteratura².

Gli stessi vengono utilizzati per determinare il coefficiente di riproduzione R_0 , alla base di molte decisioni politiche, attraverso relazioni non-lineari³.

Particolarmente utile è il confronto dei dati relativi al tasso di letalità apparente $m_r = 16\%$ e $m_r = 2,1\%$ rispettivamente per la prima e la seconda ondata. Il primo dato risulta sicuramente sovrastimato a causa dei pochissimi tamponi effettuati nella prima ondata pandemica. Il secondo dato è in linea con i dati di letteratura⁴.

Anche il tempo medio intercorrente tra un positività e un decesso, sottostimato durante la prima ondata ($t_0=6$ giorni), risente dei ritardi nell'esecuzione dei tamponi e del fatto che, nella primissime fasi della pandemia lo stesso veniva effettuato prevalentemente su persone con un stadio della malattia già avanzato.

Nella seconda ondata il dato si allinea alle due settimane citate dagli esperti confermando la bontà delle tecniche utilizzate.

Bibliografia

*La fonte dei dati è Our World in Data: <https://ourworldindata.org/coronavirus-source-data.xlsx>

¹Nelson, P.H. Biophysics and Physiological Modeling. CUP, <http://circle4.com/biophysics>

²Tian, L., Li, X., Qi, F. et al. Harnessing peak transmission around symptom onset for non-pharmaceutical intervention and containment of the COVID-19 pandemic. Nat Commun 12, 1147 (2021)

³J. Wallinga, M. Lipsitch How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. B (2007) 274, 599-604

⁴<https://www.cebm.net/covid-19/global-covid-19-case-fatality-rates/>