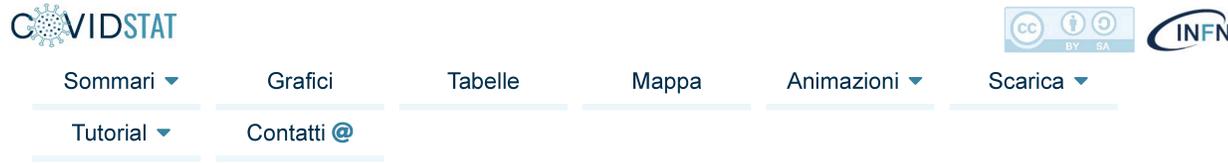


Analisi dei dati epidemiologici italiani



Il Progetto

È un progetto realizzato dal **Gruppo di Lavoro CovidStat INFN**, la cui costituzione è stata promossa in seno all'Istituto Nazionale di Fisica Nucleare con l'obiettivo di mettere a disposizione dell'Unità di Crisi Covid19 interna all'INFN una **analisi statistica dei dati forniti quotidianamente dalla Protezione Civile sulla diffusione della pandemia in Italia**. Tuttavia, si è ritenuto che gli strumenti sviluppati per questa analisi statistica possano essere di utilità anche per altre comunità scientifiche e, più in generale, possano essere di interesse pubblico.

Il sito propone quindi un'analisi strettamente statistica dei dati. Non si trattano analisi epidemiologiche, per le quali l'INFN non ha competenze specifiche. Per questo motivo il lavoro mette a disposizione un quadro aggiornato della situazione, senza fornire alcuna estrapolazione e previsione sui futuri sviluppi della pandemia.

Tutti i dati, gli istogrammi e le interpolazioni presentate in questo sito non hanno perciò alcuna pretesa di capacità predittiva perché, oltre a essere basati su modelli puramente statistici e non di tipo epidemiologico, non si conoscono gli errori sistematici, e la metodologia della raccolta dei dati è soggetta a variabilità nel tempo e sul territorio.

Il gruppo responsabile:

- Gianluca Bonifazi (UnivPM & BO)
- Luca Lista (UniNa & NA)
- Dario Menasce (MI)
- Mauro Mezzetto (PD)
- Daniele Pedrini (MI)
- Roberto Spighi (BO)

Hanno contribuito alla realizzazione del sito:

- Stefano Antonelli (CNAF)
- Fabio Bredo (PD)
- Luca Carbone (MI)
- Francesca Cuicchio (Ufficio Comunicazione)
- Mauro Dinardo (MI)
- Paolo Dini (MI)
- Rosario Esposito (NA)
- Stefano Longo (CNAF)
- Stefano Zani (CNAF)

Realizzato dall'Istituto Nazionale di Fisica Nucleare grazie al supporto delle Sezioni INFN di Napoli e di Padova.
Dati: nazionali della Protezione Civile, mondiali del JHU CSSE, dati sulle vaccinazioni da [.it Developers Italia](#). Mappa interattiva basata su [Datawrapper](#).

Nota: le informazioni riportate in questo sito sono frutto di un'analisi statistica e non hanno alcun valore predittivo.

INFN and COVID-19 Challenge, 9-10 Febbraio 2021

M. Mezzetto (INFN-PD) per il gruppo CovidStat INFN

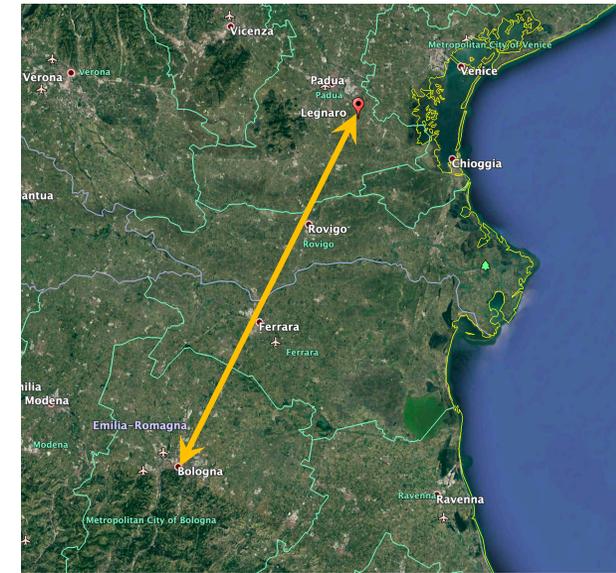
- Il sito web
- Mortalità in Italia e nel mondo
- Effective Reproduction Number R_t
- R_t calcolato con i soli sintomatici

Le risorse di calcolo necessarie (CNAF)

- Supporto nazionale da CNAF e INFN Napoli (significativo impegno di risorse HW e umane), con supporto da INFN-Pd e INFN-MiB

CNAF - LNL

- Sistema geograficamente distribuito e ridondato (due siti, CNAF Bologna – Legnaro, ~104 Km)
- Supporto per alta affidabilità (HA) e bilanciamento dei workload (funzionalità garantita nel caso di perdita di accesso a uno dei due siti).
- RTO (Recovery Time Objective) tendente a zero in caso di failure di uno dei due nodi.
- **CPU:** 16 Blade DELL M640 a due vie, dotati di CPU Intel Xeon 6132 o 6230 (8 CNAF e 8 LNL).
- 384GB di RAM, 2 connessioni FC a 16Gbps verso la SAN e 4 connessioni ethernet a 10Gbps.
- **Storage:** Huawei OceanStor 5500 v5 (totale 1Tb)
- **Sito Web:** 4 VM (covid19-01.infn.it -> covid19-04.infn.it) dotate ognuna di 4 core (Xeon 6230), 8 GB di RAM. 500 GB di storage su ogni sistema è stato gestito via GlusterFS, con 4 "brick" in replica (RAID1 distribuito).



INFN Napoli

- **Hardware:** VM 64 cores e 64GB di RAM
- **Middleware:** Jupyter HUB con supporto dei linguaggi **perl**, **C++**, **python**, **Javascript** e **R + Root (CERN)** e **Plotly**
- **Software:** script di nostra creazione per l'analisi dei dati ed il calcolo di Rt. Repository pubblico su GitHub (Baltig)

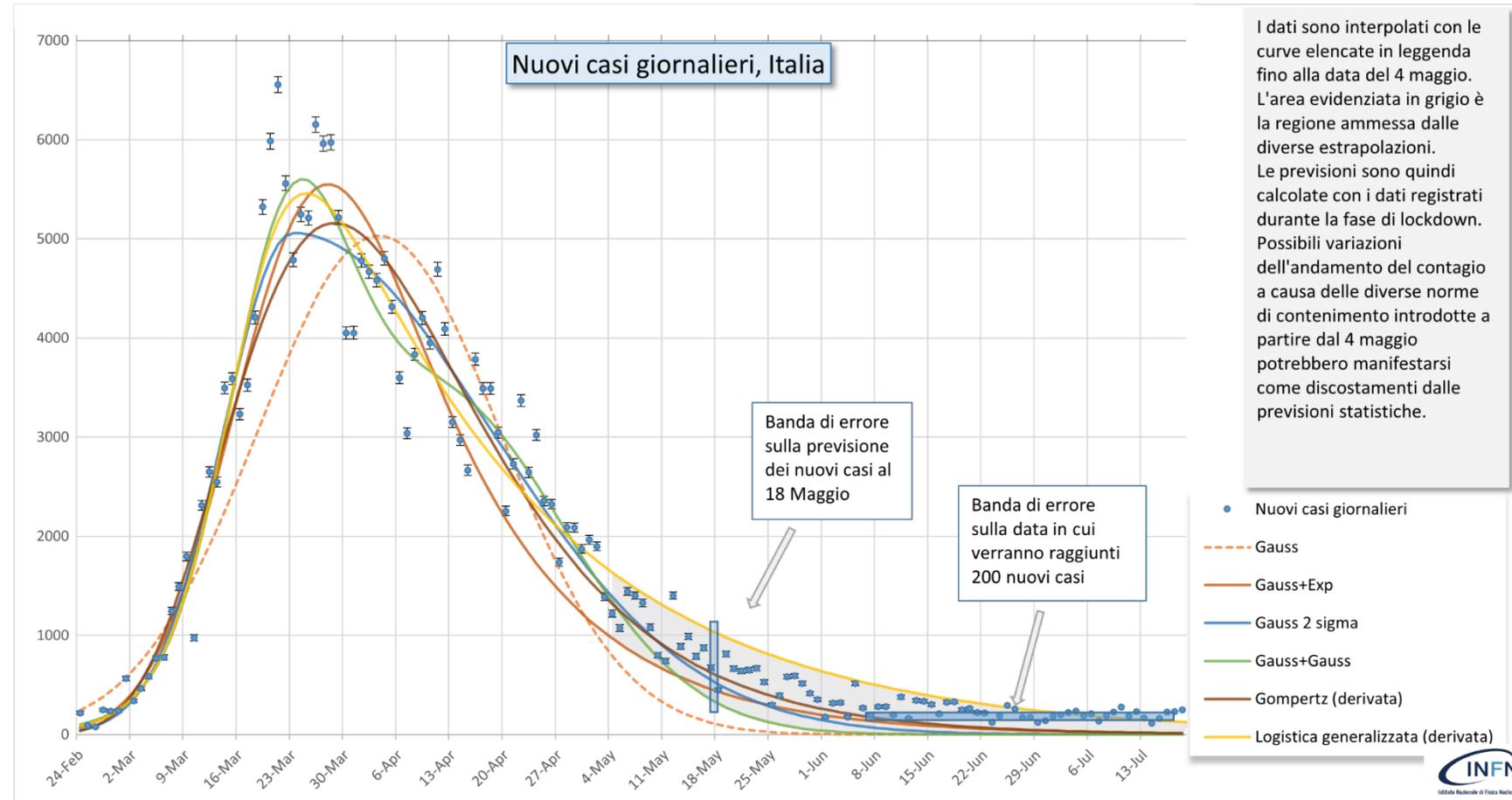
Sito Web

- Aggiornamento quotidiano **automatico** (via cron) dalle 16:40 con poll del sito GitHib della Protezione Civile
- Aggiornamento quotidiano **automatico** dalle 7:00 dei dati mondiali dalla JHU e dalle 9:00 della situazione dei vaccini

La discesa dal primo picco (i)

Durante la fase di discesa, asimmetrica rispetto alla salita, abbiamo pubblicato quotidianamente questo plot:

- Abbiamo fittato i dati del contagio fino alla data di fine lockdown (4 maggio)
- Utilizzando diverse curve asimmetriche
- Abbiamo così definito una regione di estrapolazione che tiene conto dell'incertezza del modello
- I dati del contagio sono rimasti nella banda di estrapolazione per circa due mesi
- Poi hanno smesso di decrescere

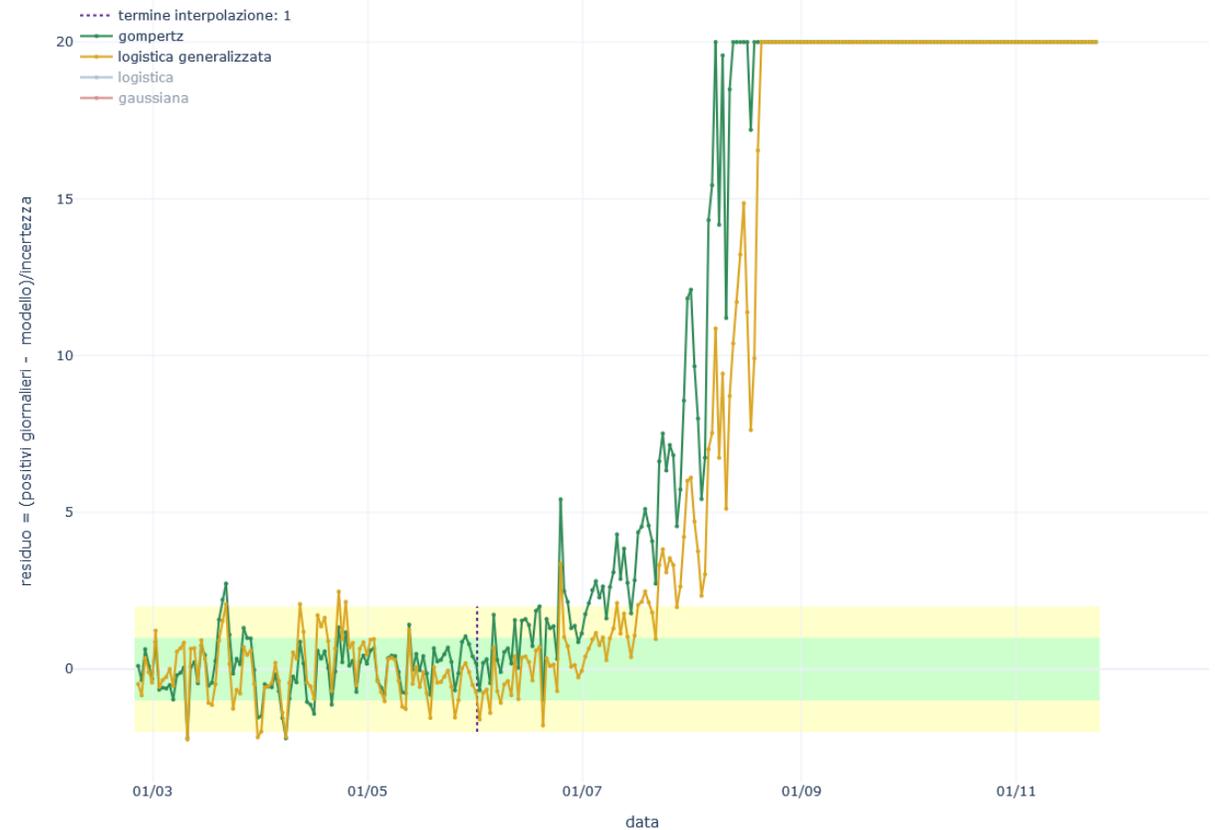


La discesa dal primo picco (ii)

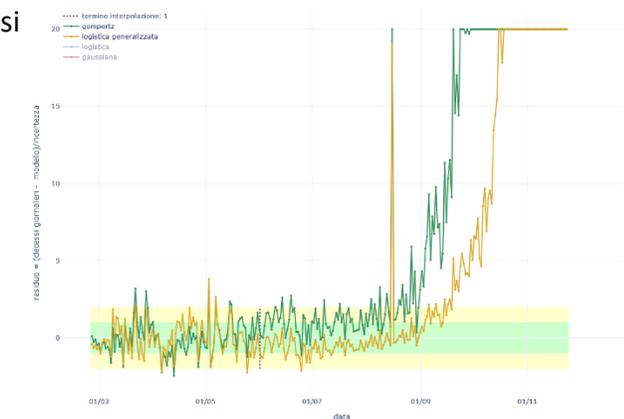
Nel modello precedente non avevamo la possibilità di quantificare la deviazione dalle previsioni.

Per cui siamo ricorsi ai residui:

- Fit ai dati del contagio fino alla fine del lockdown con Gompertz e Logistica
- Introdotto un fattore di scala sugli errori poissoniani per ricondurre la distribuzione dei residui ad una gaussiana
- Con questa rinormalizzazione è quindi possibile misurare il discostamento dei dati dalla previsione in termini di standard deviations
- È stato con questo strumento che abbiamo assistito impotenti alla ripartenza dei contagi



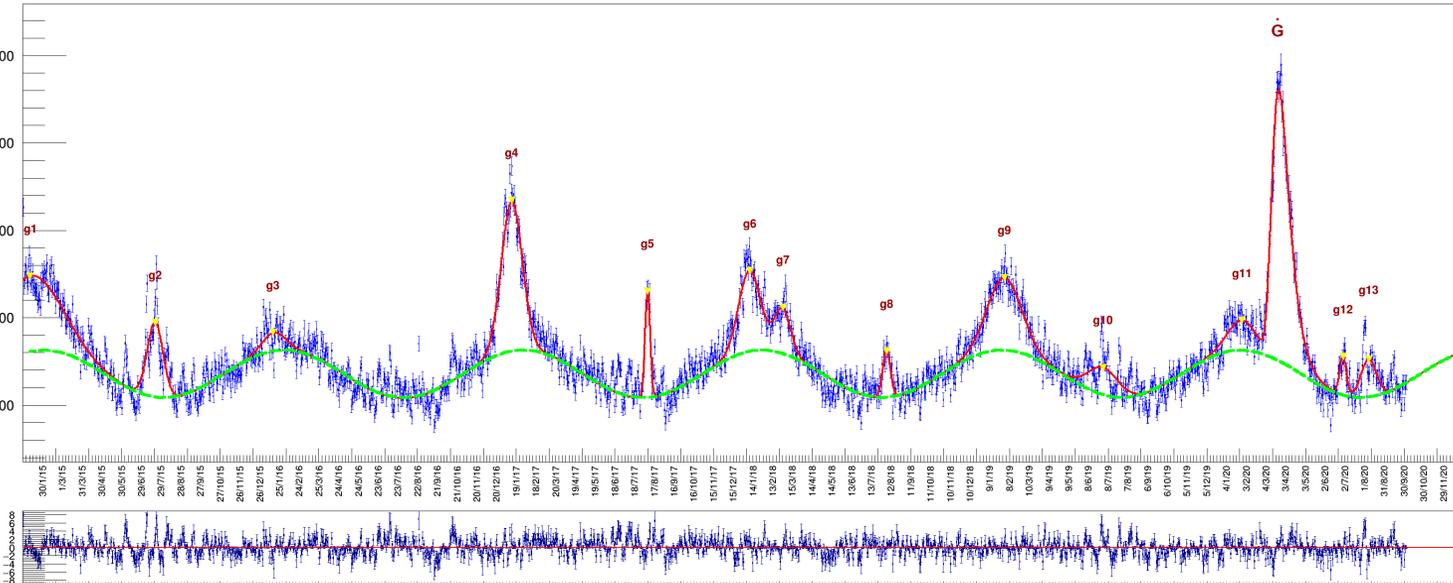
Andamento dei decessi



La mortalità al primo picco

D. Menasce et al. "A statistical analysis of death rates in Italy for the years 2015-2020 and a comparison with the casualties reported for the COVID-19 pandemic", [arXiv:2102.01629](https://arxiv.org/abs/2102.01629)

Number of daily deaths



Parametri della baseline

c	a	T (days)	φ (rad)
1678 ± 1.5	139.4 ± 2.593	364 ± 0.4	-5.27 ± 0.02

Parametri delle Gaussiane

g_i	N_i	μ_i (days)	σ_i (days)	$\mu_i - 2\sigma_i$ (date)	μ_i (date)	$\mu_i + 2\sigma_i$ (date)	Duration (days)
1	50706 ± 3092	10.9 ± 3.4	47.1 ± 2.3	9/10/2014	10/1/2015	15/4/2015	189
2	13005 ± 361	201.1 ± 0.4	12.0 ± 0.4	25/6/2015	20/7/2015	13/8/2015	49
3	4455 ± 527	381.6 ± 2.0	17.7 ± 2.1	12/12/2015	16/1/2016	21/2/2016	71
4	34015 ± 534	743.5 ± 0.2	15.9 ± 0.3	11/12/2016	12/1/2017	13/2/2017	64
5	5959 ± 210	950.1 ± 0.1	3.9 ± 0.2	30/7/2017	7/8/2017	14/8/2017	15
6	19120 ± 704	1104.9 ± 0.6	16.5 ± 0.8	6/12/2017	8/1/2018	11/2/2018	67
7	7862 ± 616	1155.0 ± 1.0	12.5 ± 1.2	3/2/2018	28/2/2018	24/3/2018	49
8	4084 ± 256	1313.3 ± 0.4	6.0 ± 0.5	24/7/2018	5/8/2018	17/8/2018	24
9	26850 ± 685	1492.5 ± 0.6	26.1 ± 0.6	10/12/2018	31/1/2019	24/3/2019	104
10	9299 ± 504	1642.3 ± 1.4	23.2 ± 1.3	15/5/2019	30/6/2019	15/8/2019	92
11	9020 ± 613	1853.3 ± 1.7	21.8 ± 1.5	14/12/2019	27/1/2020	10/3/2020	87
12	2841 ± 217	2007.5 ± 0.4	5.1 ± 0.4	19/6/2020	29/6/2020	9/7/2020	20
13	6546 ± 362	2046.3 ± 0.7	11.9 ± 0.8	14/7/2020	7/8/2020	31/8/2020	48

Parametri della Gompertz

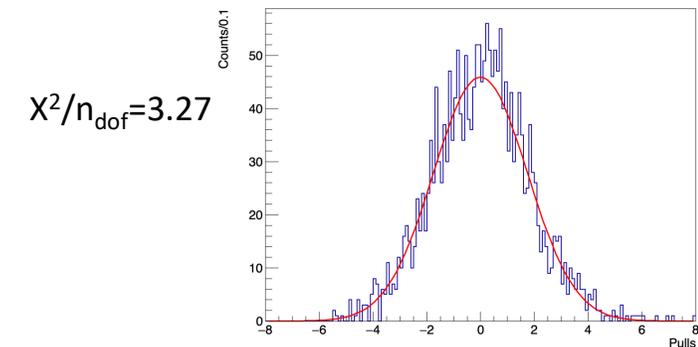
Yield	From	Peak	To	Duration (days)
54387 ± 557	20/2/2020	24/3/2020	11/5/2020	81

- Dati ISTAT sulla mortalità in Italia nel periodo 2015-2020 (*)

- Fit globale con una funzione $F(t) = s(t) + \sum_{i=1}^k g_i(t) + \dot{G}(t)$; $s(t) = c(t) + a \sin\left(\frac{2\pi t}{T} + \varphi\right)$
($g_i(t)$: Gauss; $G(t)$: Gompertz)

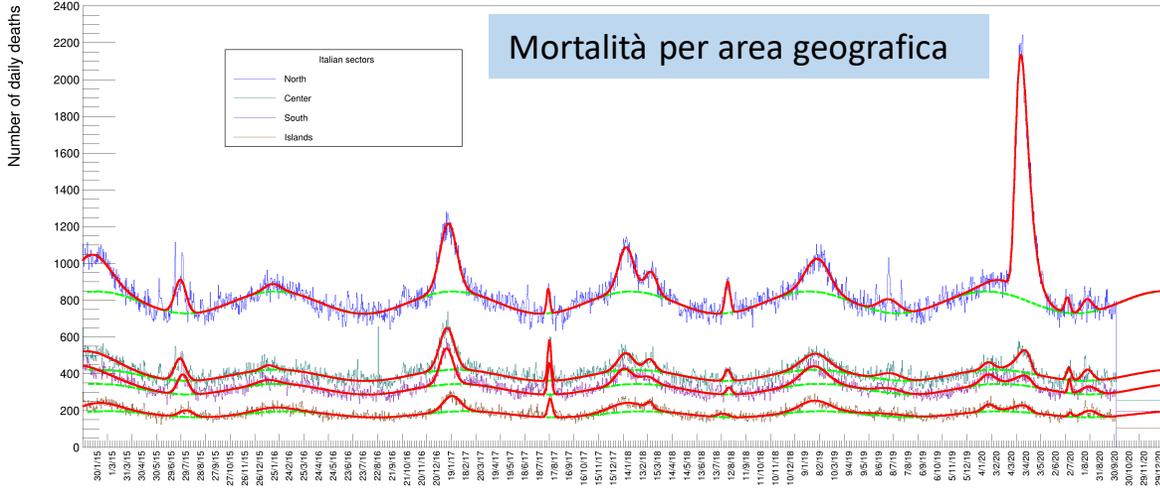
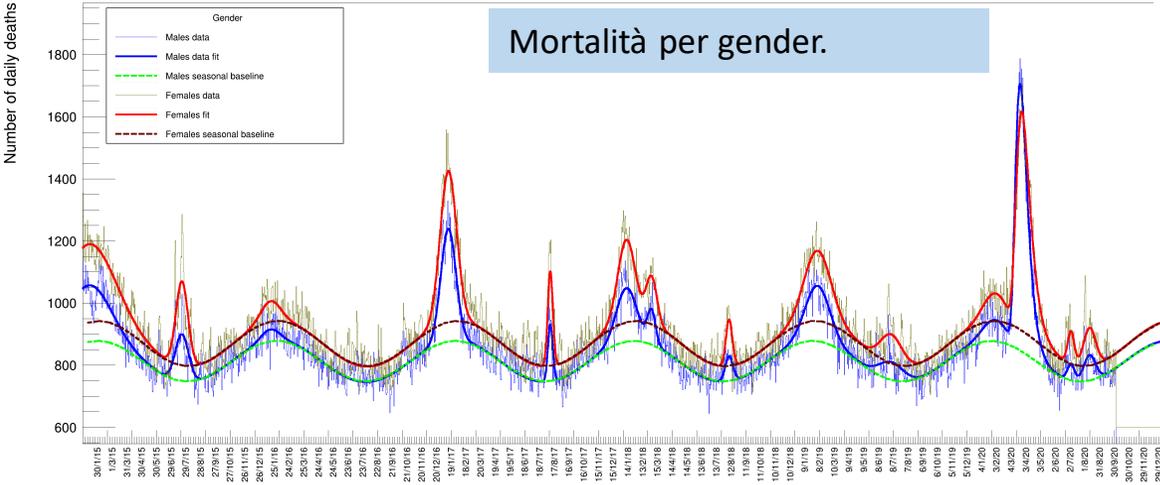
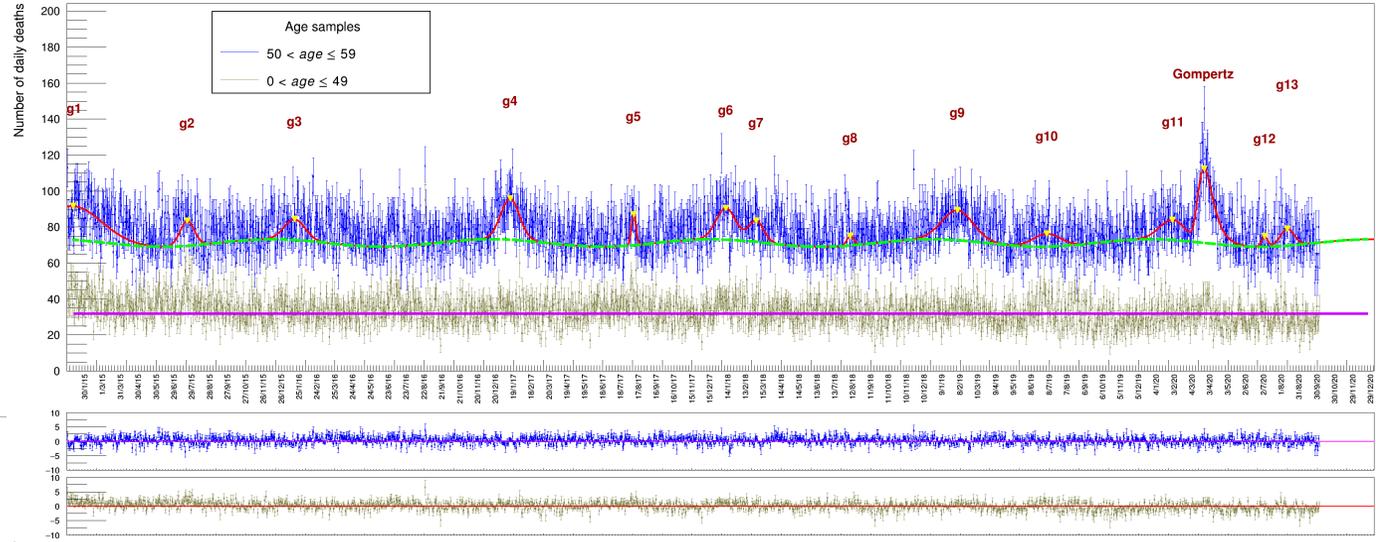
- In letteratura: a) sottrazione del fondo dei picchi calcolandolo con la media dei decessi negli ultimi 5 anni nello stesso periodo; b) sottrazione del fondo fittando la sinusoidi nei periodi in cui i picchi sono assenti (Euromomo)

Distribuzione dei pulls

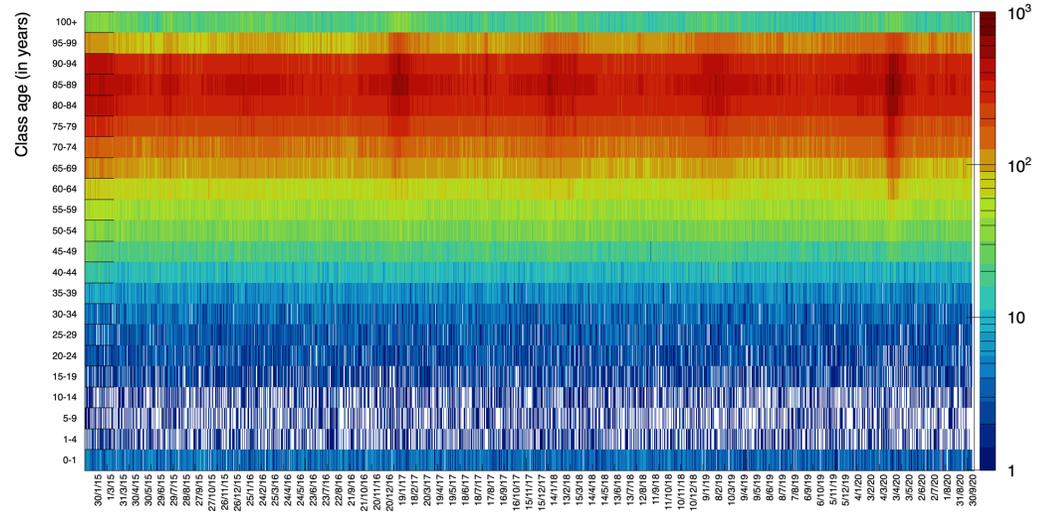


(*) https://www.istat.it/files/2020/03/Dataset-decessi-comunali-giornalieri-e-tracciato-record_dati-al-30-settembre.zip

La mortalità al primo picco (II)



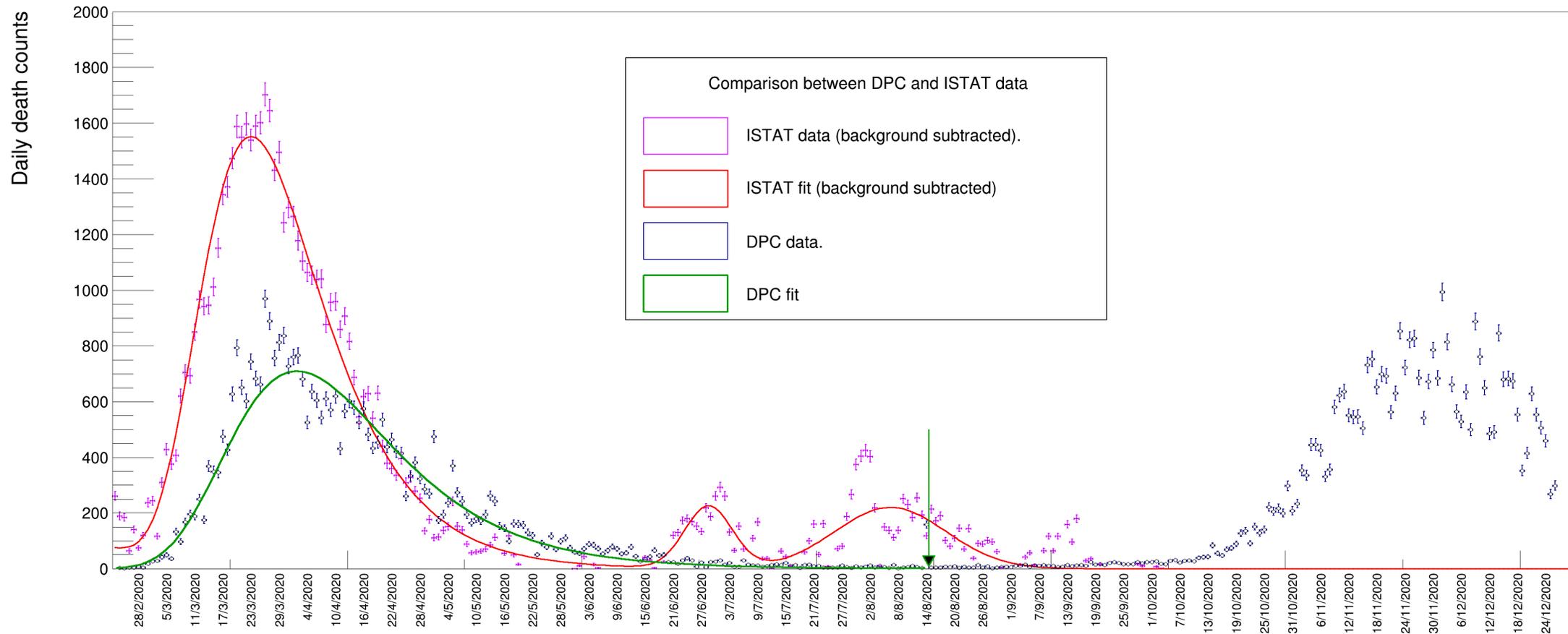
Tutte le features della distribuzione di mortalita' compaiono nel campione di eta' > 50 anni. I residui del campione 0-50 sono calcolati rispetto al fit con una semplice costante.



Scatter plot of the class age versus day of death. The dark red spots show the age and the date corresponding to highest values of deaths incidence.

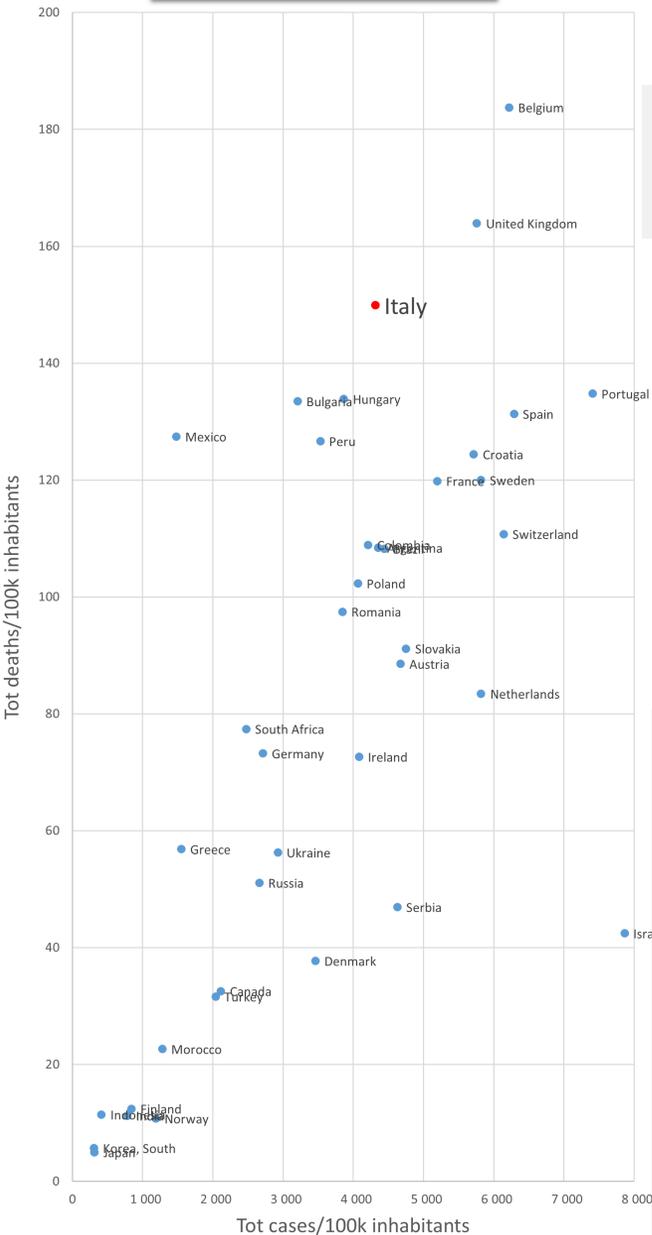
Confronto con I decessi Covid-19 riportati da DPC

- Decessi ISTAT nel periodo 28/2 -15/8, sottratto il background: 54387 +/- 557
- Decessi Covid-19 nello stesso periodo: 35488
- Va notato che ISTAT riporta i decessi alla data anagrafica, mentre DPC li riporta alla data in cui riceve comunicazione dalle Regioni (si veda ad esempio lo spike del 15/8)



Davvero la mortalità italiana è fra le peggiori al mondo?

Deaths vs Cases (per 100k inhabitants)



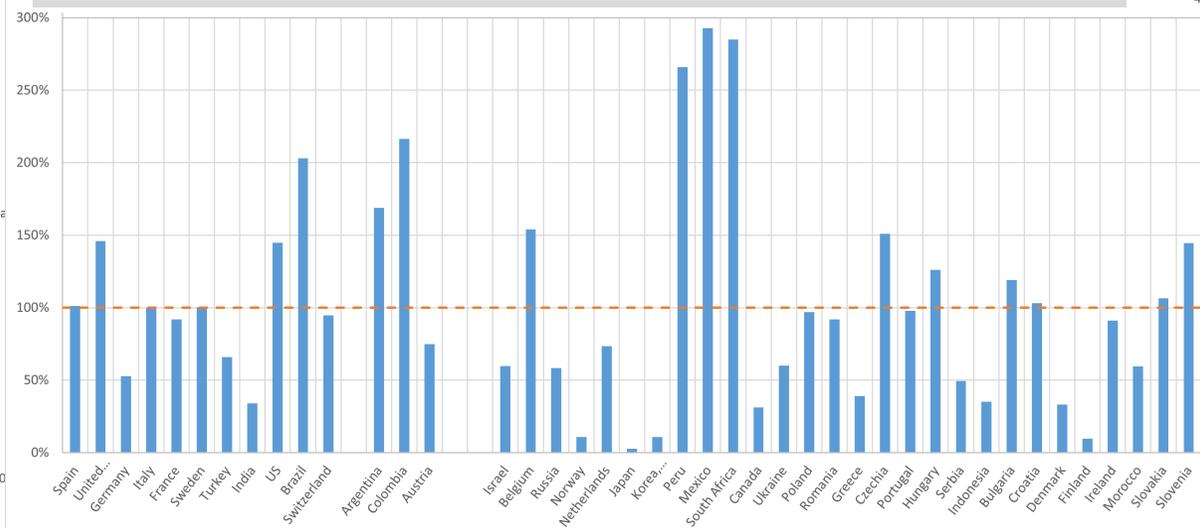
Mortalità: decessi Covid-19 normalizzati alla popolazione

1) Apparentemente sì: mortalità vs casi totali normalizzati alla popolazione per 42 paesi nel mondo

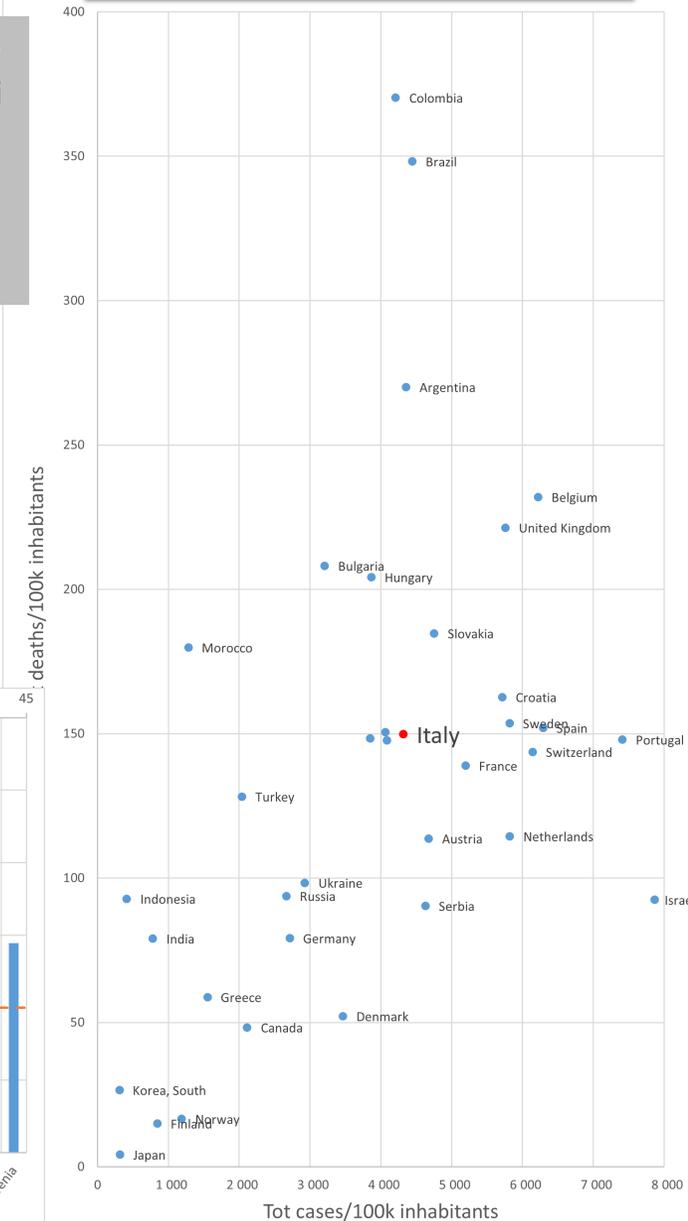
3) Oppure si può rinormalizzare la popolazione per fasce di ogni nazione su quella italiana, assumendo che il rapporto di mortalità nelle varie fasce sia uguale a quello italiano

4) La via maestra sarebbe confrontare la mortalità per fasce di età delle varie nazioni (e naturalmente definire decessi per Covid-19 nello stesso modo in tutto il mondo)

2) Nota di L. Baglini e P. Spagnolo: va normalizzata per le diverse fasce di età: rapporto fra il numero di decessi reale e quello calcolato con le fasce di età della nazione considerata moltiplicate per la mortalità per fasce dell'Italia

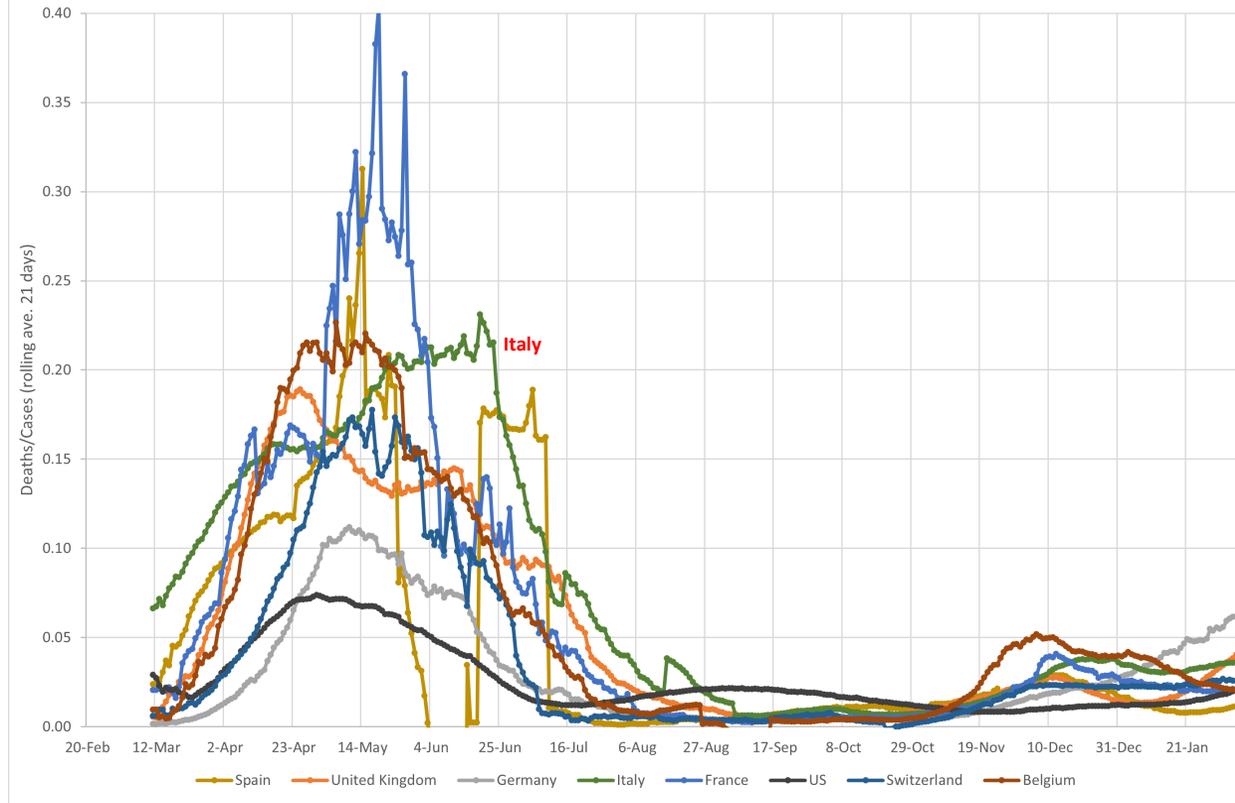


Deaths vs Cases (deaths normalized to the Italian distribution of population)



Case Fatality Rate (CFR)

Case Fatality Rate

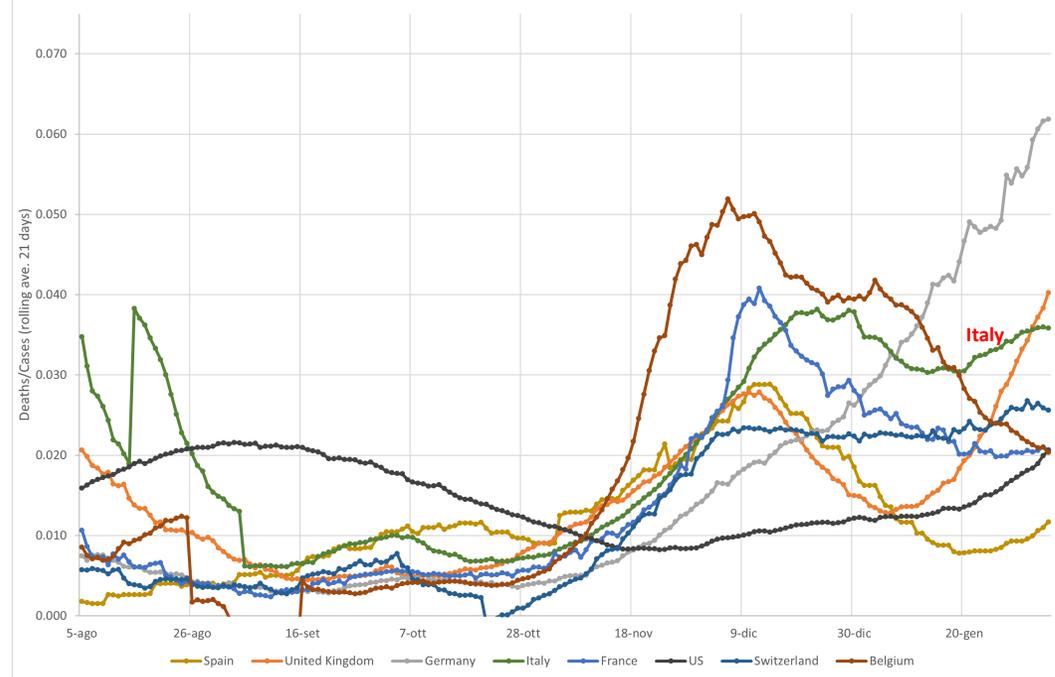


Decessi normalizzati sul numero di contagi.

Fortemente modulata dal denominatore, che dipende dal numero di tamponi. È quasi una misura dell'efficienza di tracing

L'andamento di Germania e UK nell'ultimo periodo è comunque fortemente preoccupante

Case Fatality Rate

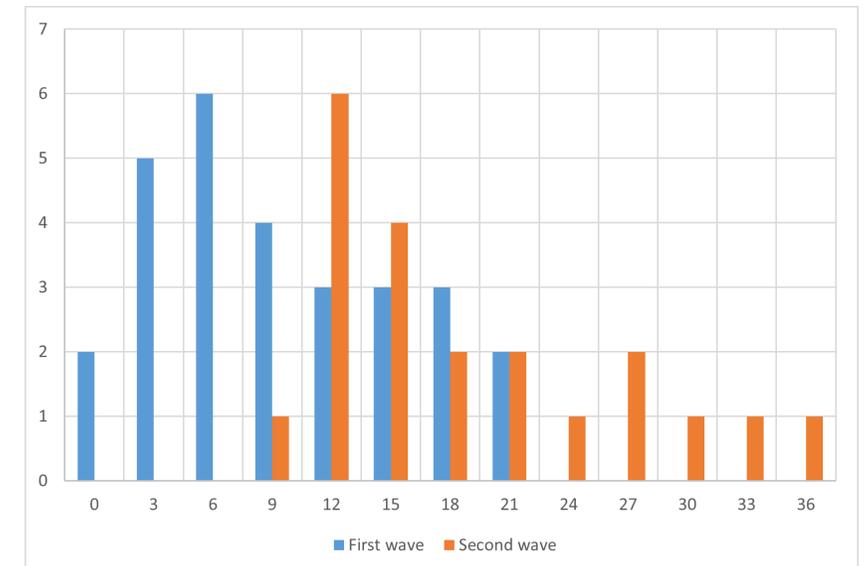
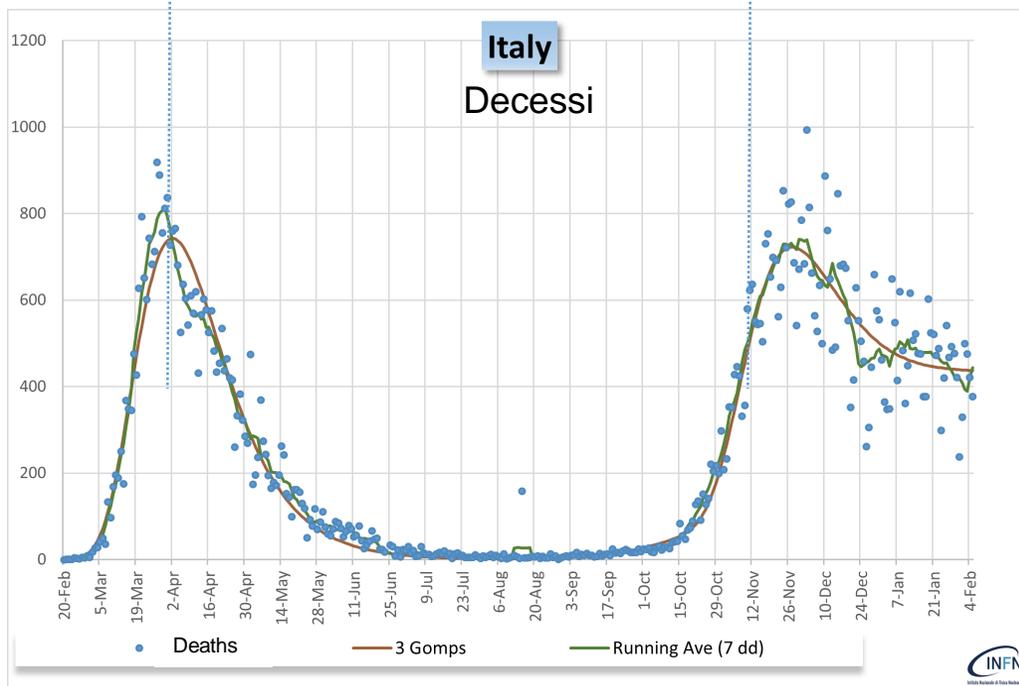
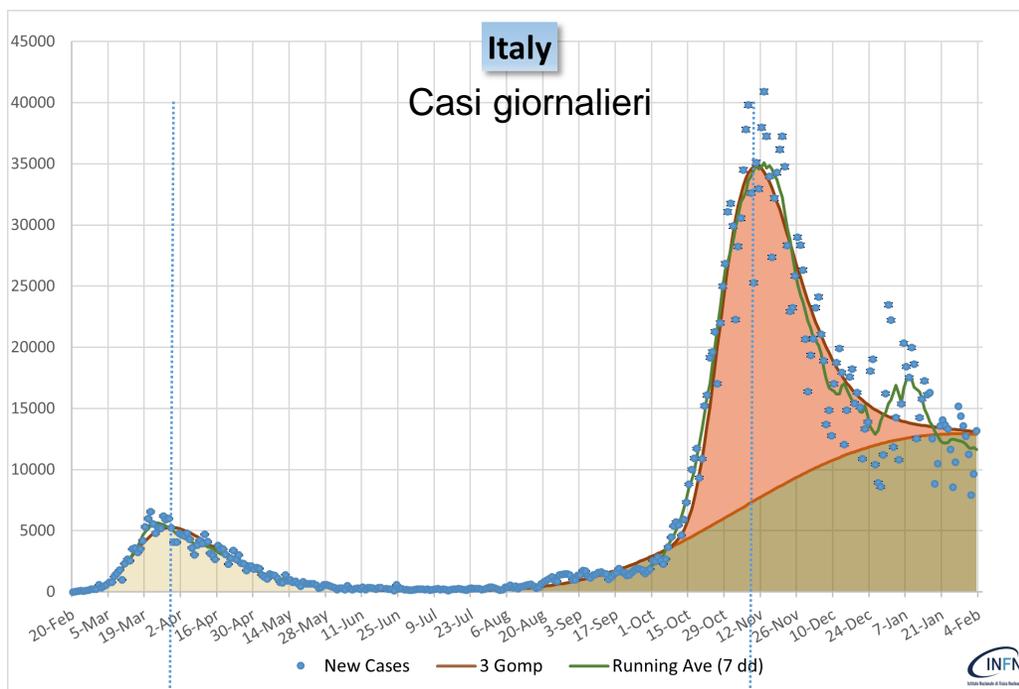


Un'ultima osservazione sulla mortalità'

Fit ai nuovi casi giornalieri e ai decessi in Italia utilizzando la somma di tre Gompertz

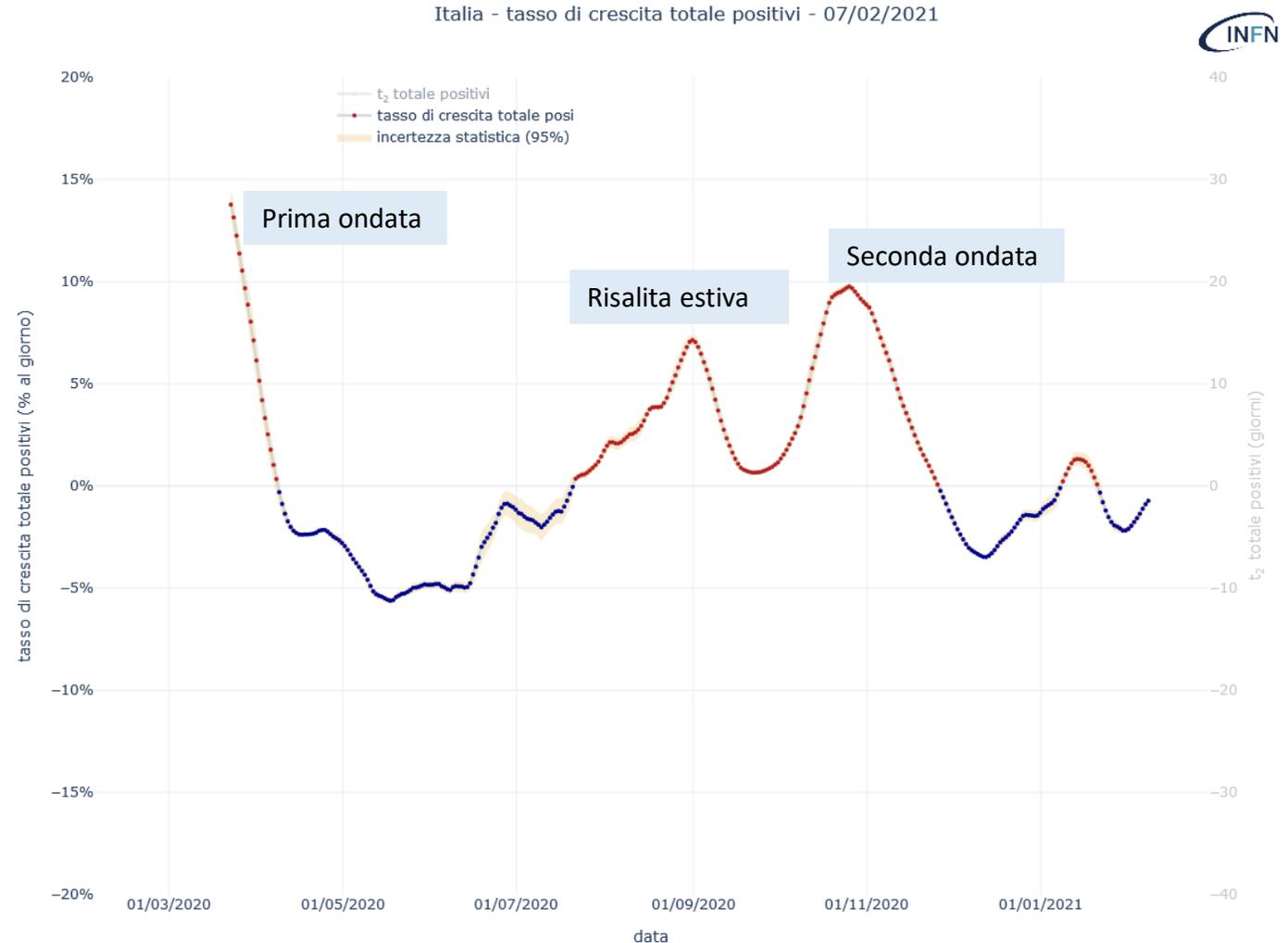
Al primo picco il massimo dei decessi è avvenuto con appena 4.0 giorni di ritardo rispetto al massimo dei nuovi casi
Nel secondo picco il ritardo dei decessi è stato di 15.6 giorni.
È patologico dell'Italia?

Distribuzione di Δt al primo e al secondo picco per le 40 nazioni al mondo con più casi di Covid-19.



La seconda ondata

- Le curve che abbiamo utilizzato per descrivere la fase di discesa, non sono in grado di fittare i dati in modo convincente nella fase di salita (cominciano a dare un fit ragionevole solo in prossimità del picco)
- L'indicatore migliore dello sviluppo del contagio in questa fase è il tempo di raddoppio, quindi un fit esponenziale agli ultimi 14 (o 7) giorni
- Nello storico del contagio il tempo di raddoppio presenta diversi asintoti (al raggiungimento del picco il tempo di raddoppio tende a + infinito, per ripartire da - infinito subito dopo il picco)
- Per questo motivo preferiamo riportare l'inverso del tempo di raddoppio, il sito comunque restituisce sia il tempo di raddoppio che il suo inverso passando il mouse sui punti della curva
- Siamo consapevoli che questa curva tende a risultare indigeribile ad un utente occasionale del sito.



Effective Reproduction Number R_t

R_t , effective reproduction number, rappresenta il numero di contagi prodotti in media da ogni contagiato al tempo "t"

METODI SIRD

$$S(t) = S(t-1) - \frac{\alpha}{N} S(t-1) I(t-1)$$

$$I(t) = I(t-1) + \frac{\alpha}{N} S(t-1) I(t-1) - \beta I(t-1) - \gamma I(t-1)$$

$$R(t) = R(t-1) + \beta I(t-1)$$

$$D(t) = D(t-1) + \gamma I(t-1)$$

$$R_{SIRD}(t) = \frac{\alpha}{\beta + \gamma}$$

S(t): Susceptibles

I(t): Infected (not total cases)

R(t): Recovered

D(t): Deaths

N: Entire population

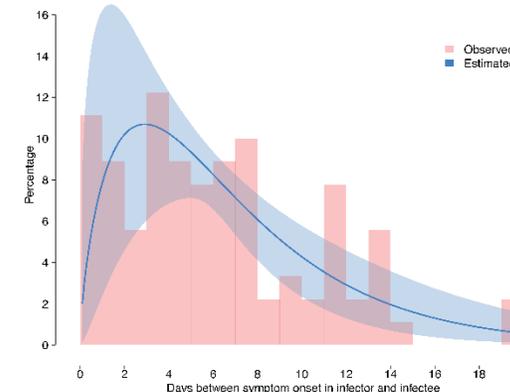
Assuming $S(t)/N=1$

$$E(I_t) = R_t \sum_{s=1}^t I_{t-s} w_s$$

w_s : generation time: time interval between infector-infected, normalized to unity

Vari metodi proposti in letteratura per estrarre R_t da questa equazione

Il primo metodo che abbiamo implementato. Segue molto bene il contagio ma i suoi valori assoluti non sono in accordo con i valori di ISS



Generation function calcolata su 90 casi in Lombardia all'inizio del contagio. è una gamma function con

- Media = 6.6 +/- 1.88 gg
- Std = 4.88 +/- 1.03 gg

arXiv:2003.09320 "The early phase of the COVID-19 outbreak in Lombardy, Italy"

Gamma function

$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}$$

k: shape
 θ : scale
 $\Gamma(k)=(k-1)!$

Media=k θ
 Std= $k\sqrt{\theta}$

Metodi per il calcolo di R_t

Dalla pagina di documentazione del sito CovidStat

Wallinga-Teunis (2004)

R_t viene stimato direttamente dall'andamento del numero di contagiati I_t , stimato dal numero di positivi giornalieri, assumendo di conoscere la distribuzione di probabilità w_s ^[1].

Viene usata una procedura statistica basata sulla probabilità che due contagiati, i e j , con sintomi a tempi $t_i > t_j$, siano una possibile coppia contagiante-contagiato, normalizzata dalla probabilità che il caso i sia stato contagiato da ogni altro caso k . In formule:

$$p_{ij} = w_{t_i-t_j} / \sum_{i \neq k} w_{t_i-t_k}.$$

Da questo si ottiene $R_j = \sum_i p_{ij}$, e la stima di R_t si ottiene come la media di tutti i valori R_j dei contagi al giorno t .

Per questo motivo, e per coerenza con gli altri algoritmi, riferiamo il valore calcolato non al primo giorno dell'intervallo considerato, ma al settimo. Questo metodo non può essere quindi usato per stimare R_t negli ultimi giorni di osservazione.

Per il calcolo di questo algoritmo utilizziamo il pacchetto open source `EpiEstim`^[2]. Il solo contributo considerato nel calcolo dell'incertezza è quello statistico.

Bibliografia - metodo di Wallinga-Teunis

[1] Wallinga, Teunis (2004), *American Journal of Epidemiology*, Volume 160, Issue 6, p. 509–516.

[2] `EpiEstim`: Estimate Time Varying Reproduction Numbers from Epidemic Curves.

Cori et al. (2013)

R_t viene stimato dal rapporto del numero di persone contagiate I_t al tempo t rispetto al numero di possibili contagi al tempo t , dato da $\sum_{s=1}^t I_{t-s} w_s$ ^[1].

Per calcolare R_t in questo modo viene introdotta una procedura Bayesiana. Va notato che in questo approccio R_t viene calcolato in base al numero di contagiati registrato nei giorni precedenti e non a quelli successivi. Per questo viene anche definito R_t istantaneo.

Per calcolare R_t con questa procedura è necessario definire la distribuzione w_s . Noi utilizziamo la distribuzione pubblicata da Cereda et al.^[2], che è la stessa utilizzata dall'Istituto Superiore di Sanità nelle sue stime.

Per il calcolo di questo algoritmo utilizziamo il pacchetto open source `EpiEstim`^[3]. La banda di incertezza, riportata come intervallo di credibilità Bayesiano, tiene conto dell'incertezza statistica di Poisson e degli errori con cui è conosciuta w_s .

Bibliografia - metodo di Cori et al.

[1] Cori et al. (2013), *American Journal of Epidemiology*, Volume 178, Issue 9, p. 1505–1512.

[2] Cereda et al., *arXiv:2003.09320* (2020)

[3] `EpiEstim`: Estimate Time Varying Reproduction Numbers from Epidemic Curves.

Metodi per il calcolo di R_t

Dalla pagina di documentazione del sito CovidStat

Bettencourt Ribeiro (2008)

Il metodo è simile a quello di Cori *et al.*, ma al posto della funzione w_s viene considerato un intervallo di tempo fisso. R_t viene stimato con un Markov Chain Monte Carlo e non con calcoli analitici^[1]. La banda di incertezza, riportata come intervallo di credibilità Bayesiano.

Per il calcolo di questo algoritmo utilizziamo il metodo proposto da Kevin Systrom^[2] e implementato nel sito Rt COVID-19 Italia^[3]

Bibliografia - metodo di Bettencourt-Ribeiro

[1] Bettencourt, Ribeiro (2008) PLoS ONE, Volume 3, Issue 5, e2185.

[2] Kevin Systrom, *The Metric We Need to Manage COVID-19*.

[3] G. Bonifazi, Rt COVID-19 Italia, <https://rt-italy.live/>.

RKI

L'istituto RKI, Robert Koch Institut^[1] calcola R_t come il rapporto tra la somma del numero di contagiati negli ultimi 4 giorni e la somma del numero dei contagiati nei 4 giorni precedenti. La motivazione di questa stima di R_t è descritta in un articolo dell'RKI^[2] disponibile in tedesco.

Esistono pacchetti software per fare questo calcolo ma, vista la semplicità dell'algoritmo, abbiamo preferito implementarlo con nostro codice che abbiamo [validato confrontando i nostri risultati con quelli pubblicati in base ai dati tedeschi](#).

È necessario applicare una procedura di *smoothing* dell'andamento dei numeri dei contagiati giornalieri per ridurre l'oscillazione alla modulazione settimanale. Infatti, nel fine settimana tipicamente il numero di tamponi analizzati è più basso che negli altri giorni, di conseguenza il numero di contagiati rilevato è anch'esso più basso. Abbiamo pertanto applicato un filtro di Savitzky-Golay^[3] con una finestra di 15 giorni ed una polinomiale di 3° grado.

L'incertezza riportata al 68% CL è solo quella statistica che tiene conto della distribuzione poissoniana del numero di conteggi.

Bibliografia - metodo dell'RKI

[1] Robert Koch Institut, Berlin, Germany.

[2] Robert Koch Institut, *Erläuterung der Schätzung der zeitlich variierenden Reproduktionszahl* (2020).

[3] Savitzky, Golay, *Anal. Chem.*, 36, 8, 1627–1639 (1964).

Un metodo “esatto” per il calcolo di R_t

L. Lista et al. “A simplified estimate of the Effective Reproduction Number R_t using its relation with the doubling time and application to Italian COVID-19 data”, arXiv:2012.05194

Confrontando la formula di R_t : $E(I_t) = R_t \sum_{s=1}^t I_{t-s} w_s$ con uno sviluppo esponenziale $I_t = Ae^{\lambda t}$

si ottiene che per un generation time costante $w_s = g$ risulta $R_t = e^{g\lambda}$

Se w_s non è una costante si può dimostrare che può essere sostituito da $g_{eff} = g - \lambda \frac{\sigma^2}{2}$

dove λ e σ sono la media e la std della distribuzione.

Se infine w_s è una funzione gamma (come sempre assunto in letteratura), si dimostra che

$$R_t = (1 + \lambda\theta)^k$$

Dove k e θ sono i parametri di shape e scala della funzione gamma.

Questa formula originale, esatta nel caso di generation time descritto da una funzione gamma, è quella che utilizziamo di default nel sito CovidStat

Se R_t viene utilizzato solamente per descrivere l'andamento del contagio (R_t ha naturalmente molti altri utilizzi in epidemiologia), allora è più ragionevole utilizzare il tasso di crescita λ che non soffre delle incertezze legate alla misura del generation time

L'algoritmo è decisamente semplice da codificare.

Questo è il codice Python per il suo calcolo.

```
import numpy as np
import scipy.stats

# INPUT
SHAPE = 1.87
SCALE = 3.57
INTERVAL = 14
NDAYS_COV = 6.6
NDAYS_COV_ERR = 1.88
CUMULATIVE_CASES = [13, 33, 58, ..., 2636738]

x = list(range(INTERVAL))
y = np.log(np.convolve(CUMULATIVE_CASES))

slope, error = []
for i in range(INTERVAL, len(CUMULATIVE_CASES)):
    slope[i], error[i] = scipy.stats.linregress(x, y[i-INTERVAL:i])

rt_covidstat = lambda x: (1.0 + x * SCALE) ** SHAPE

# OUTPUT
rt = rt_covidstat(slope)
rt_err = rt * np.sqrt(((slope * NDAYS_COV_ERR)**2 + (error * NDAYS_COV)**2))
rt_cl_high_68 = rt + rt_err + (((rt_err / rt)**2) * rt / 2)
rt_cl_low_68 = rt - rt_err + (((rt_err / rt)**2) * rt / 2)
rt_cl_high_95 = rt + (2 * rt_err) + (2 * (((rt_err / rt)**2) * rt))
rt_cl_low_95 = rt - (2 * rt_err) + (2 * (((rt_err / rt)**2) * rt))
```

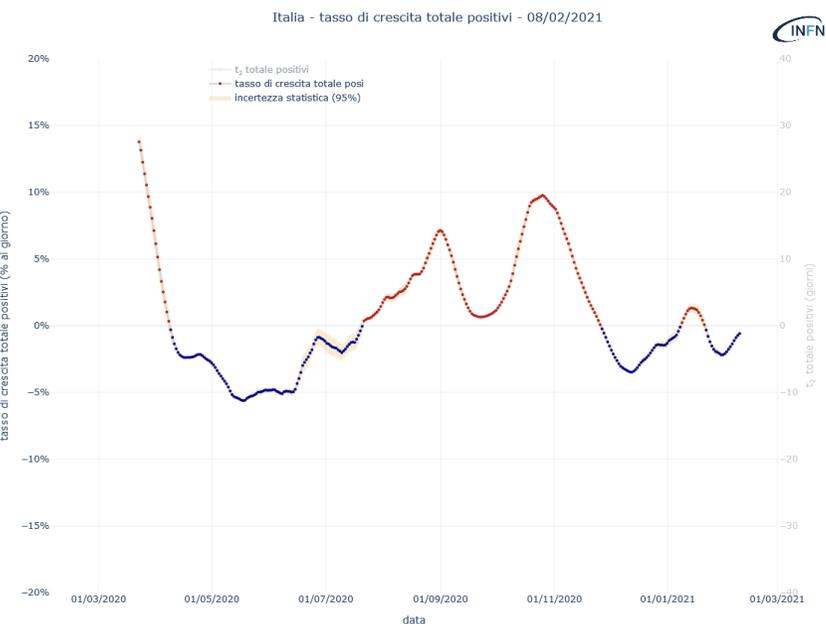
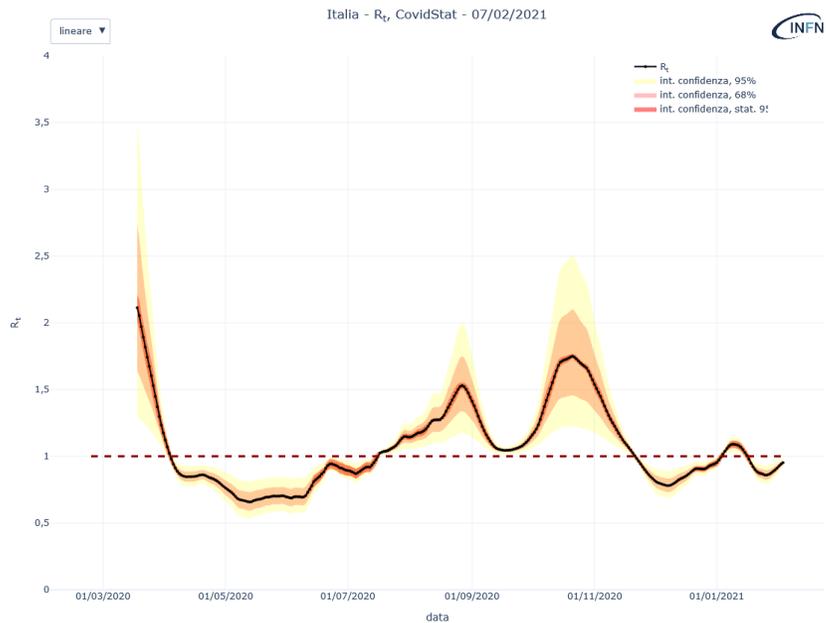
R_t nel sito CovidStat

Nel sito CovidStat pubblichiamo quotidianamente il valore di R_t calcolato con i 5 algoritmi descritti precedentemente per l'Italia, tutte le regioni italiane e per una selezione delle nazioni del mondo. I 5 algoritmi forniscono previsioni in buon accordo fra loro.

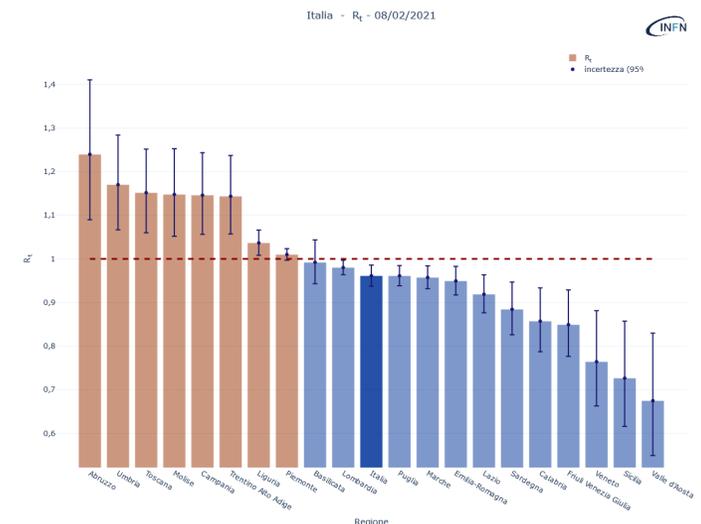
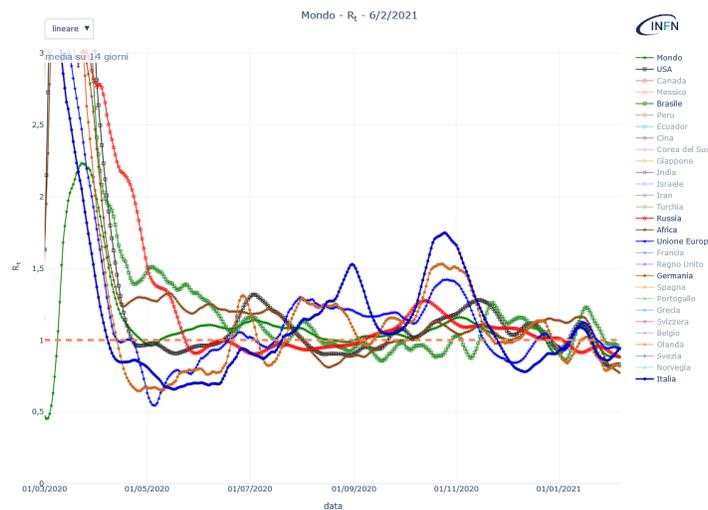


R_t e tasso di crescita λ

A riprova di quanto affermato, dalla relazione $R_t = (1 + \lambda\theta)^k$ l'andamento di R_t è identico a quello del tasso di crescita λ senza gli errori introdotti dal generation time w_s



Dal sito CovidStat: andamento di R_t in vari paesi del mondo (sx) e R_t odierno nelle regioni italiane (dx)



R_t calcolato sui positivi sintomatici

M. Mezzetto et al. "A study on the possible merits of using symptomatic cases to trace the development of the COVID-19 pandemic", arXiv:2101.01414

Il valore ufficiale di R_t per l'Italia viene calcolato dall'ISS per i soli positivi sintomatici, riportando la data di comparsa dei sintomi. Viene utilizzato l'algoritmo Cori et al.

I dati dei sintomatici non sono riportati dalla Protezione Civile, ma dal 6/12/2020 sono riportati, solamente per il totale dell'Italia, dall'ISS in <https://www.epicentro.iss.it/coronavirus/sars-cov-2-dashboard>

È necessario quindi tarare il campione che noi utilizziamo per il calcolo di R_t con quello ufficiale. In linea di principio il campione dei sintomatici presenta degli evidenti vantaggi e svantaggi

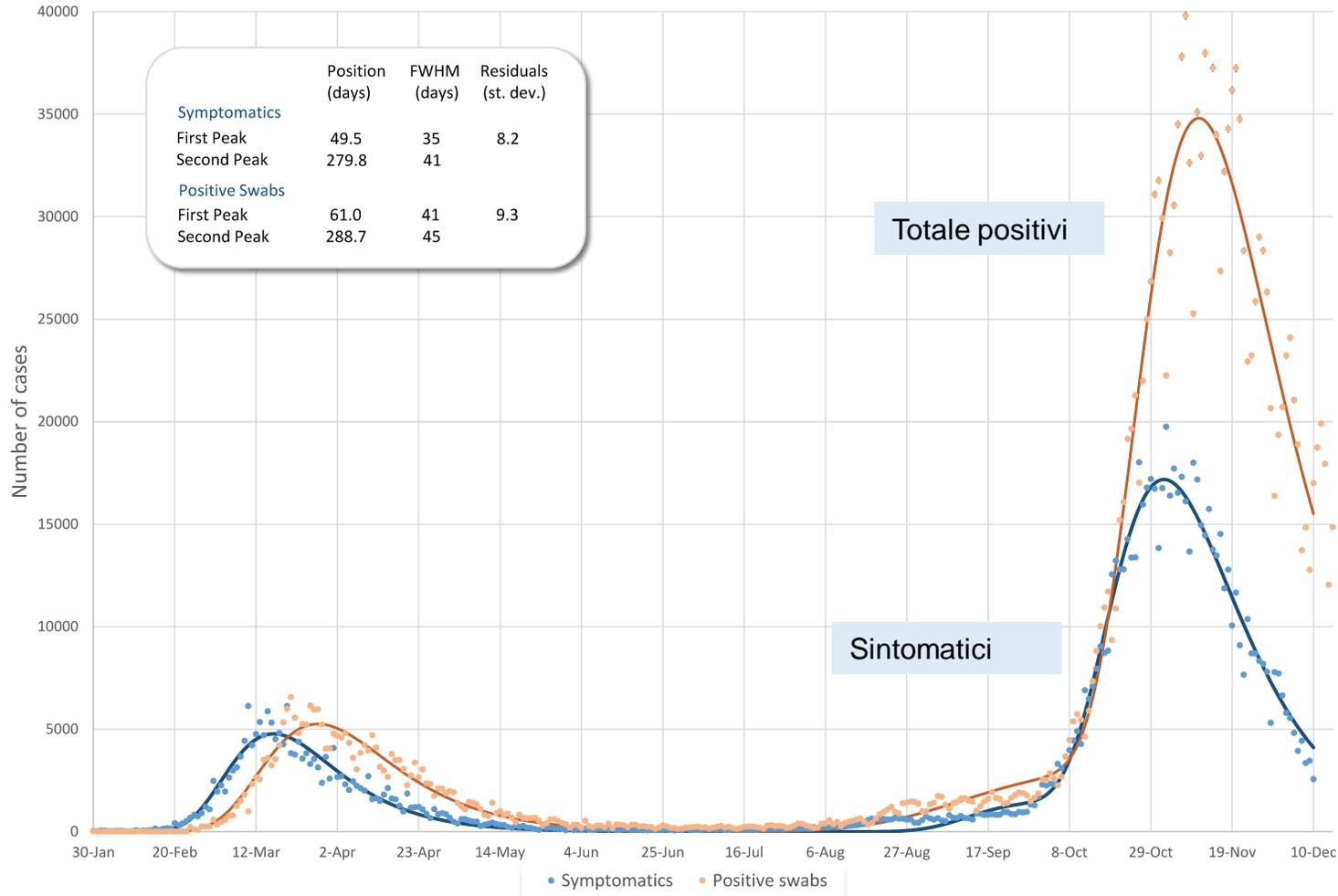
PRO

- La data riportata è più coerente con la definizione di generation time
- Campione meno sensibile alle variazioni del numero di tamponi
- Potenzialmente meno fluttuazioni dovute alla procedura di tracciamento.

CONTRO

- Il campione statistico viene ridotto
- Anche gli asintomatici sono infettivi
- Possibile bias se la frazione di sintomatici varia nel tempo (i.e. varia l'età media dei contagiati)
- La raccolta dei dati è molto più complessa.
- Difficilmente confrontabile con il resto del mondo.

Andamento dei casi totali e sintomatici



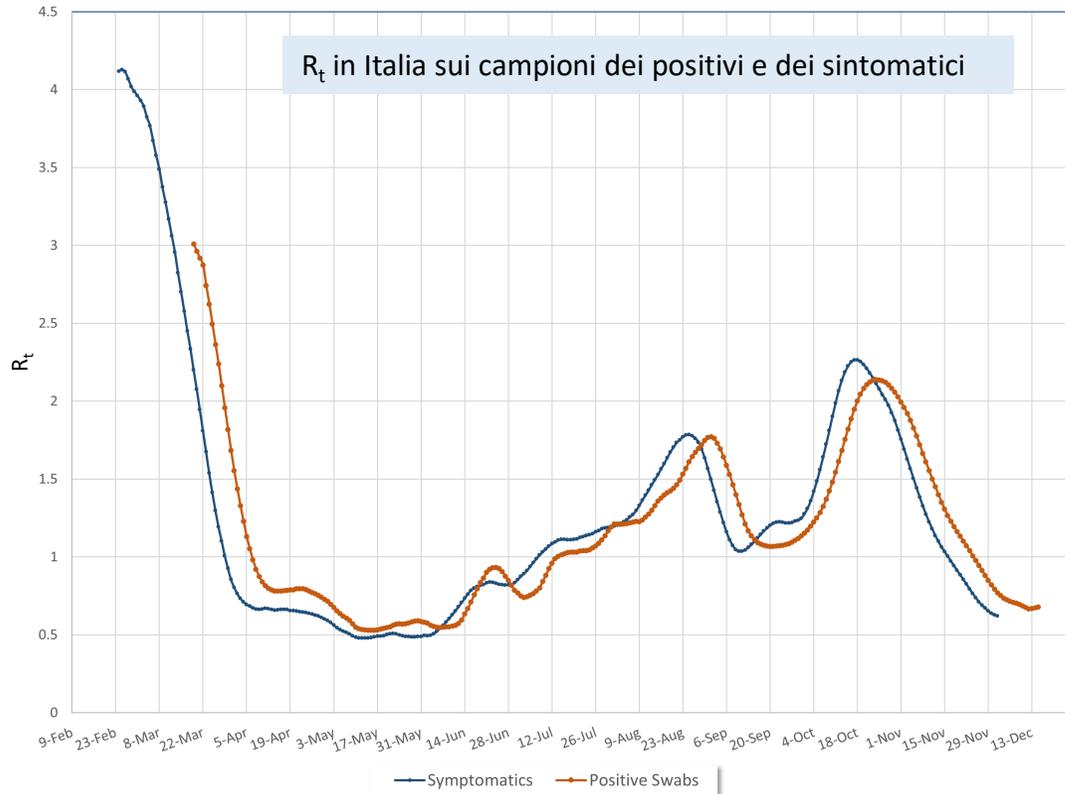
Fit ai due campioni di dati con la somma di tre Gompertz.

Posizione dei picchi: il campione dei tamponi positivi è ritardato di 9 giorni

Analisi dei residui: le due distribuzioni hanno fluttuazioni molto simili

FWHM: il campione dei tamponi positivi aggiunge una dispersione di circa 8 giorni, dovuta probabilmente alle procedure di raccolta tamponi e loro processamento

Andamento di R_t e frazione di asintomatici



Andamento di R_t calcolato con l'algoritmo Covidstat sul totale dei casi positivi e sui casi sintomatici.

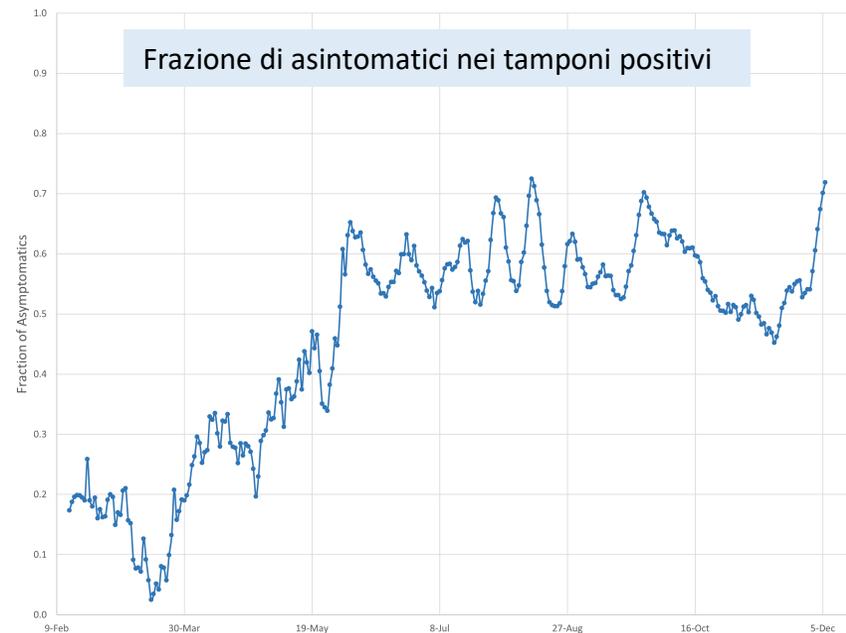
Andamento assolutamente analogo, di nuovo con un ritardo di 9 giorni. Identici risultati con R_t calcolato con gli altri quattro algoritmi presentati precedentemente.

Da notare che i casi sintomatici devono essere completamente raccolti per stabilizzarsi (oggi non ho tutti i tamponi di chi ha manifestato i sintomi pochi giorni fa). Questo impone che il calcolo dell'ISS non possa utilizzare gli ultimi 14 giorni di dati a disposizione, mentre con i dati dei tamponi positivi R_t viene calcolato fino all'ultimo giorno successivo.

In definitiva il campione dei tamponi positivi risulta sufficientemente robusto da poter essere utilizzato per il calcolo di R_t , questo ci permette di calcolare con lo stesso algoritmo il valore di R_t dalle provincie italiane alle nazioni del mondo.

Anticipando di 9 giorni la data dei campioni positivi è possibile “mettere in tempo” le due curve e calcolare la frazione di asintomatici nei tamponi positivi durante l'andamento del contagio.

È evidente che al massimo della prima ondata praticamente tutti i tamponi erano utilizzati per tracciare i sintomatici, e anche durante la seconda ondata al picco del contagio la capacità di tracciare gli asintomatici è significativamente diminuita.



Conclusioni

- Abbiamo cercato di mettere a disposizione della comunità degli strumenti che consentano una migliore comprensione dell'andamento del contagio.
- Le competenze sviluppate nell'ambito della fisica delle alte energie possono essere riutilizzate nell'analisi dei dati del contagio.
- È di fondamentale importanza mettere a disposizione della comunità scientifica il maggior numero possibile di informazioni riguardo all'andamento del contagio, con un approccio genuinamente open-source. Questo permetterebbe di effettuare analisi indipendenti e innovative su molti aspetti della pandemia.
- Nello stesso modo sarebbe fondamentale un miglior coordinamento a livello mondiale per uniformare la qualità dei dati raccolti.
- Nella speranza di produrre strumenti adeguati per poter affrontare le prossime pandemie (if any) con maggiore efficienza.

Per la prima volta dall'inizio della pandemia i contagi nel mondo stanno calando.

