

[Sommary](#) ▾[Grafici](#)[Tabelle](#)[Mappa](#)[Animazioni](#) ▾[Scarica](#) ▾[Tutorial](#) ▾[Contatti @](#)

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.

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.

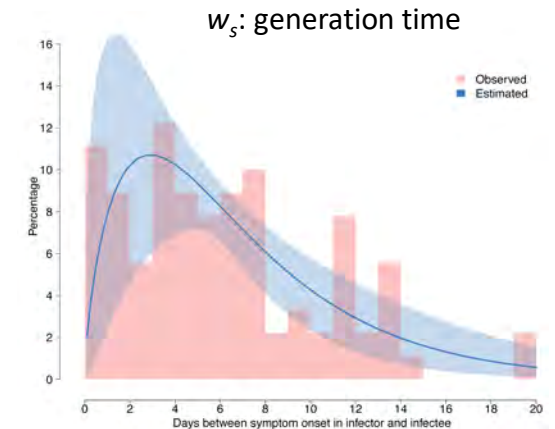
Se infine w_s è una funzione gamma (come sempre assunto in letteratura), si dimostra che

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

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 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

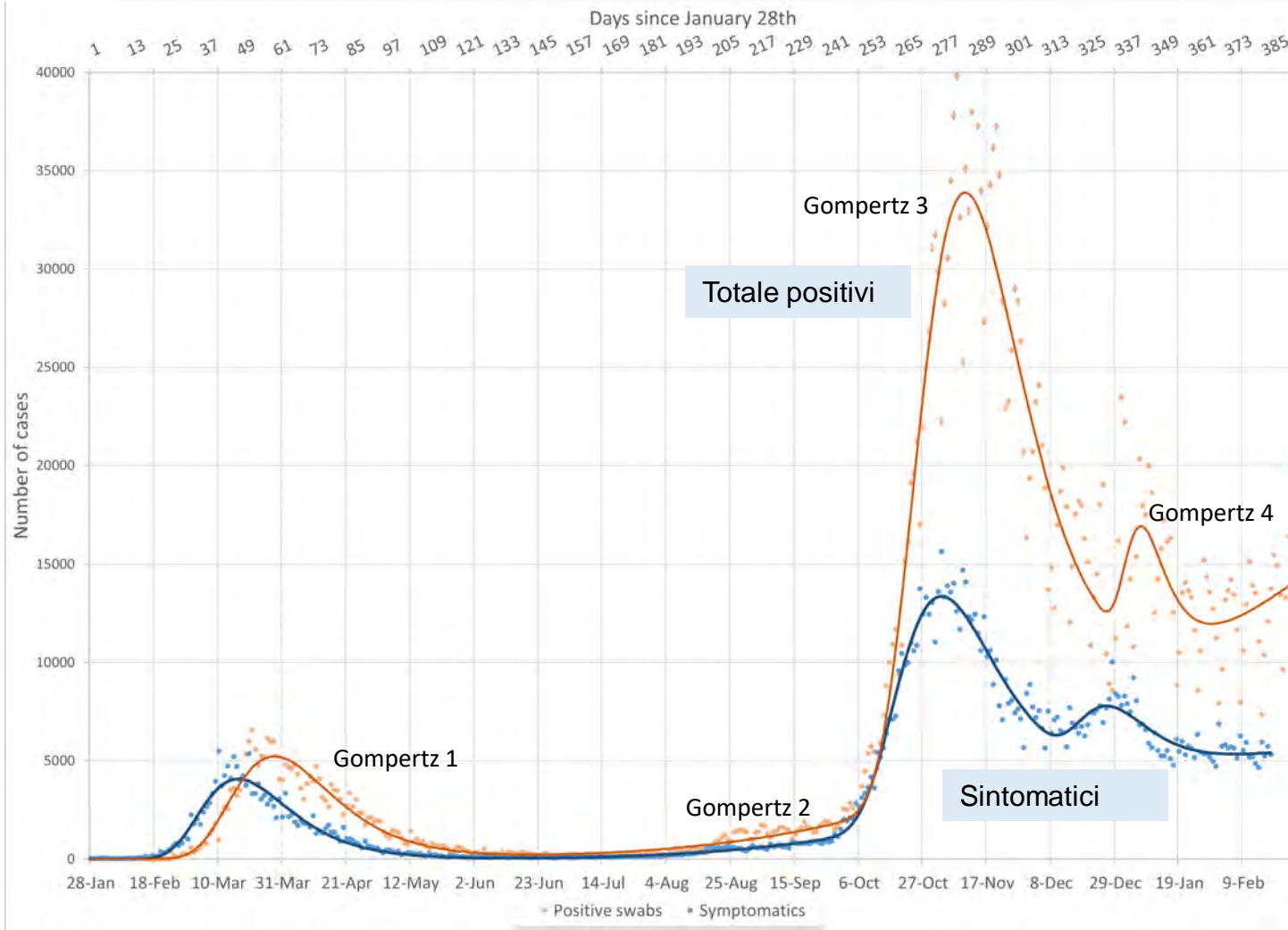
PRO

- 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 4 Gompertz.

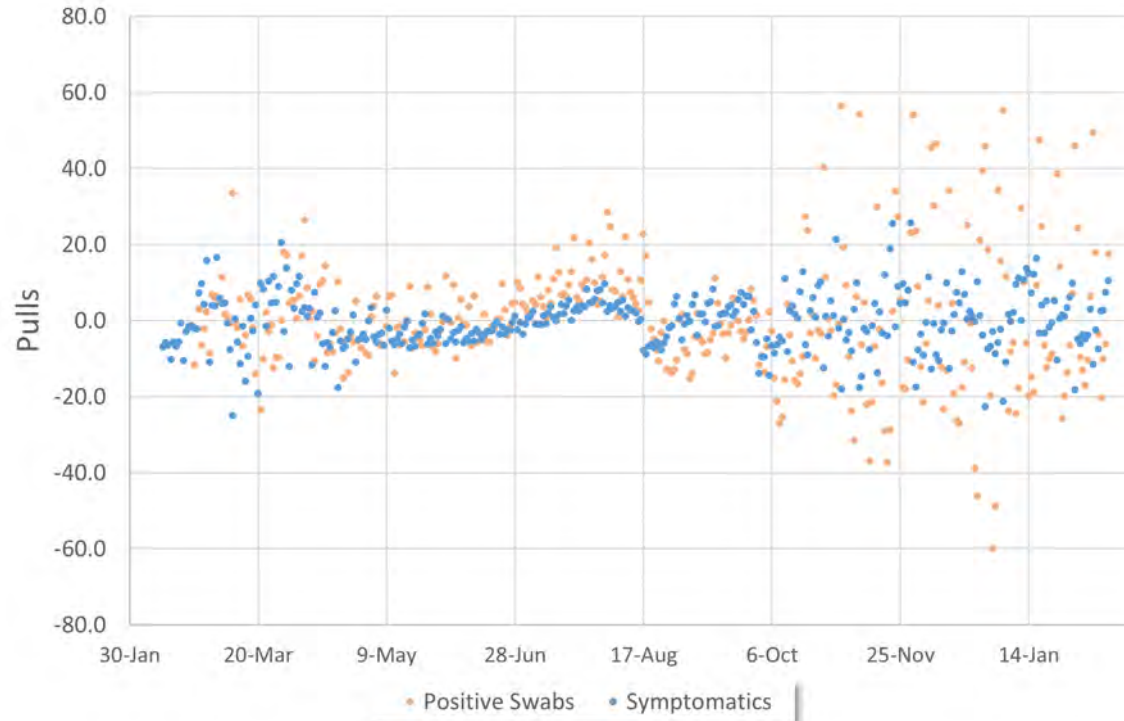
Posizione dei picchi: il campione dei tamponi positivi è ritardato di 8 giorni (al secondo picco)

Analisi dei residui: le due distribuzioni hanno fluttuazioni molto simili al primo picco, al secondo picco il totale positivi peggiora la dispersione

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

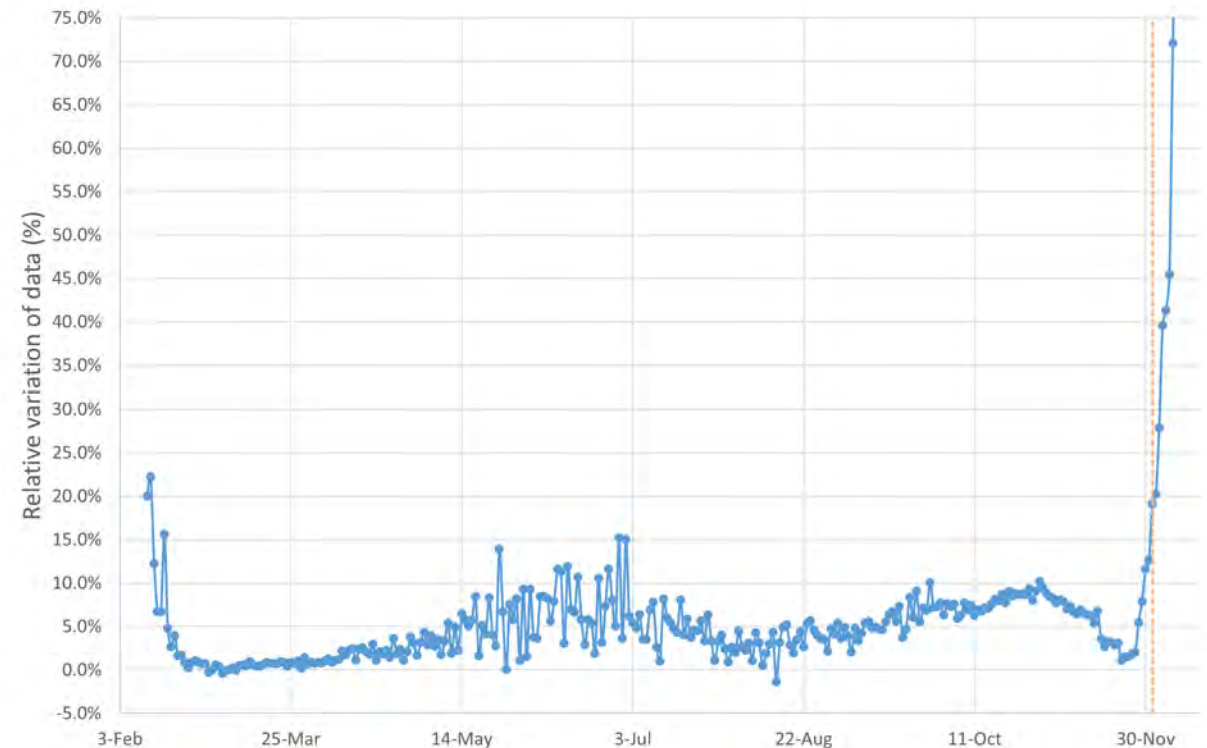
		Position (days)	FWHM (days)	Pulls (st. dev.)
Symptomatics	First Peak	49.4 ± 0.1	35.0 ± 0.1	7.2
	Second Peak	279.2 ± 0.1	51.0 ± 0.2	7.1
Positive swabs	First Peak	61.2 ± 0.3	39.4 ± 0.5	8.4
	Second Peak	287.1 ± 0.2	49.9 ± 0.3	16.2

Due importanti osservazioni

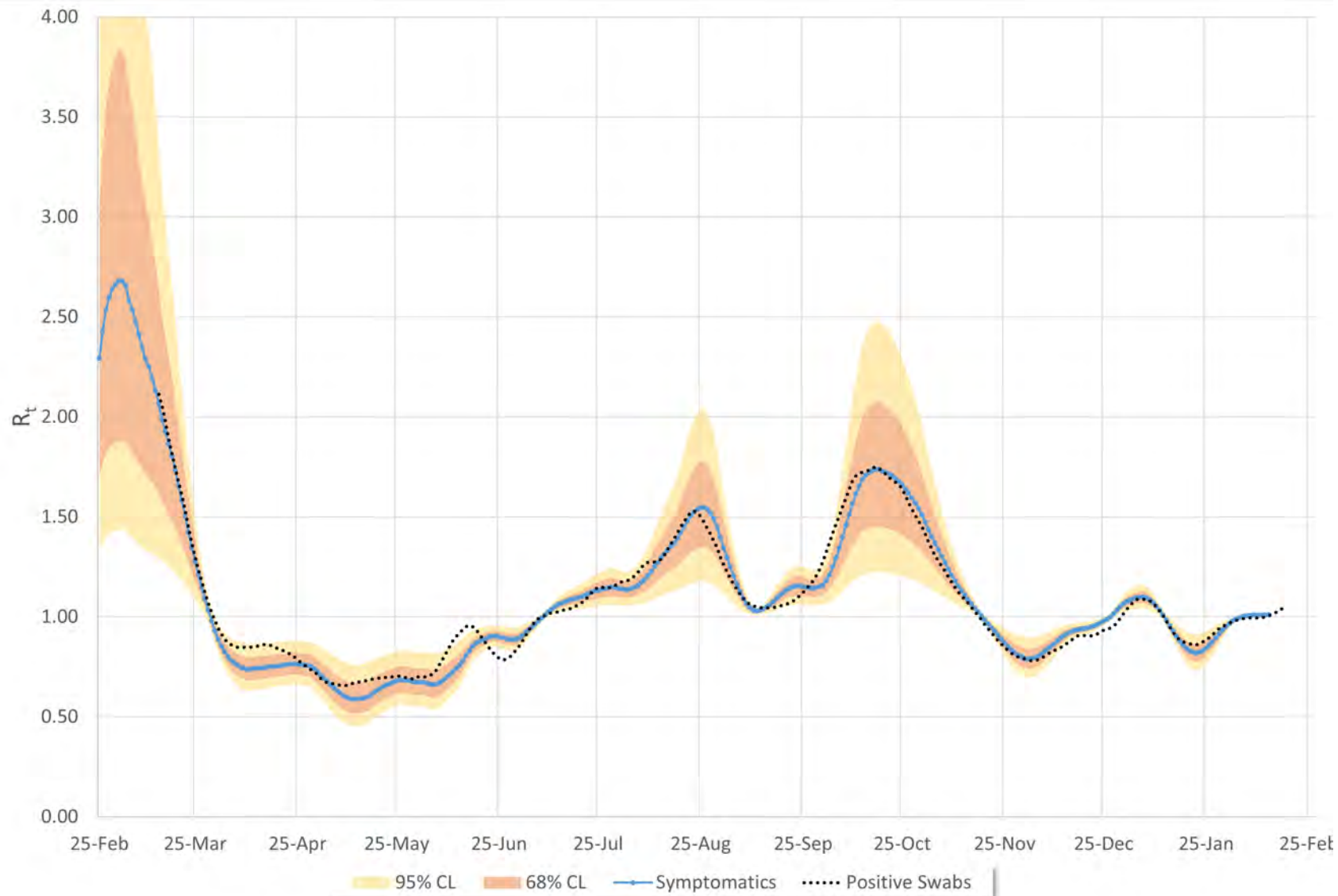


La distribuzione dei pulls della distribuzione dei nuovi casi positivi mostra che da quando sono stati introdotti i tamponi antigenici, la dispersione dei dati dei tamponi positivi e' nettamente peggiorata

Confrontando i dati dei sintomatici raccolti dall'ISS il 15Dec20 e il 1Mar21, si nota che 14 giorni non sono del tutto sufficienti per stabilizzare il campione (linea tratteggiata arancione) e che i riconteggi interessano tutto il periodo di presa dati, con variazioni fino al 15%. Questo vuol dire che l' R_t dell'ISS e' soggetto a continue revisioni



R_t calcolato sui positivi sintomatici e sui tamponi positivi



Calcolato con l'algoritmo Covidstat

I tamponi positivi sono anticipati di 8 giorni come misurato dai dati

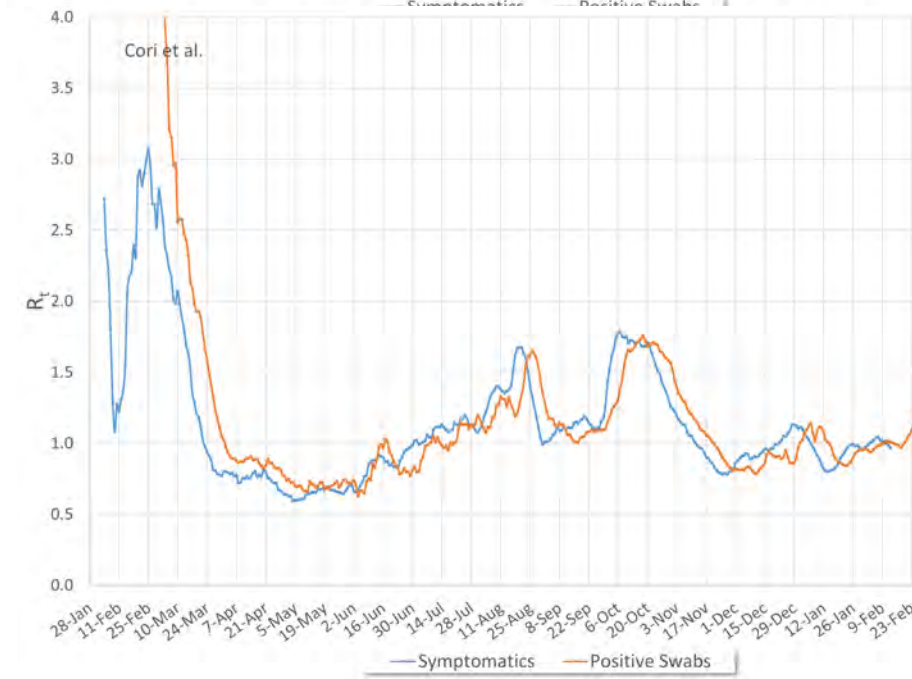
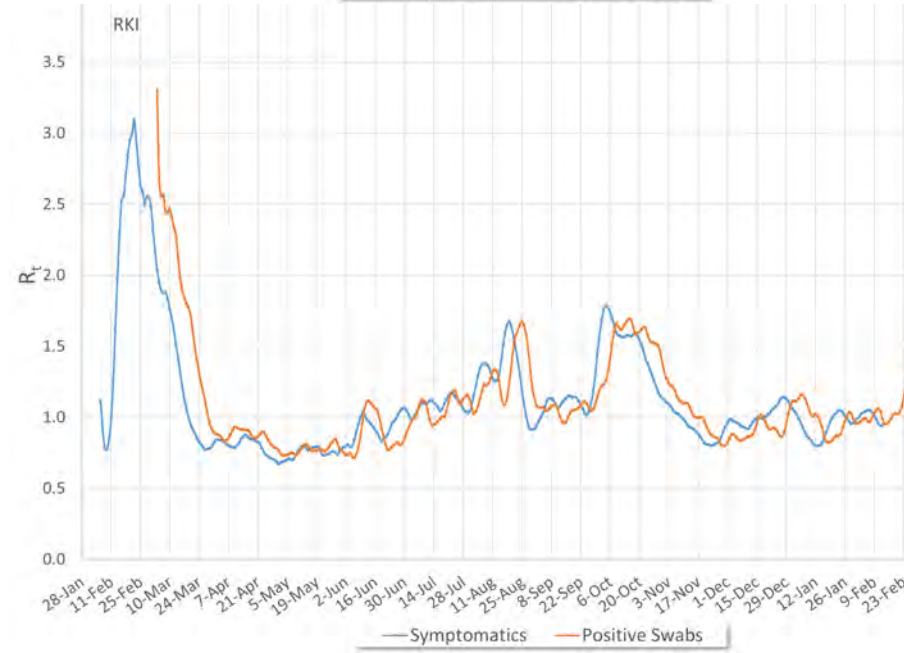
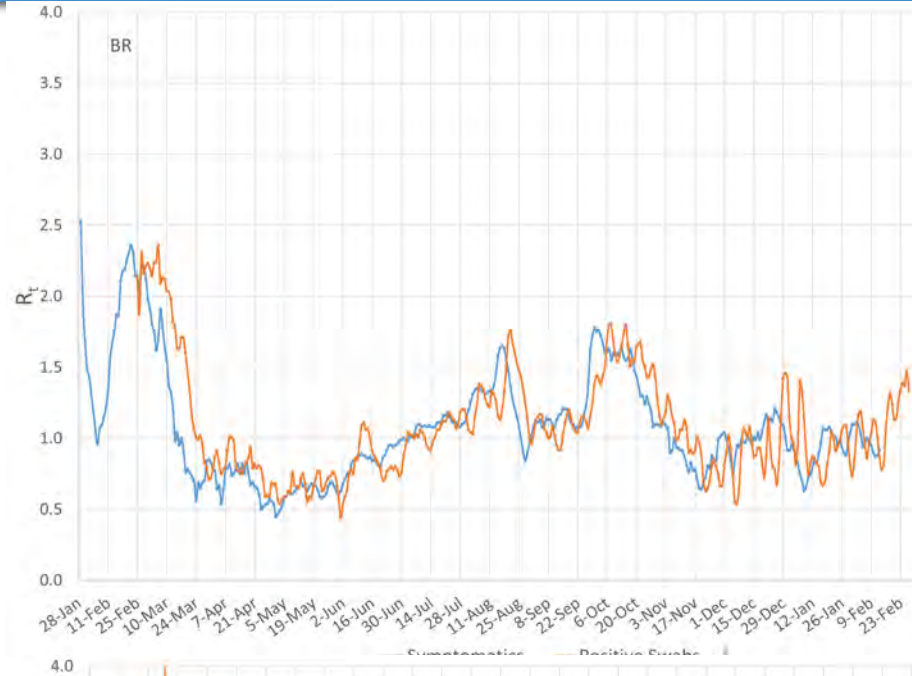
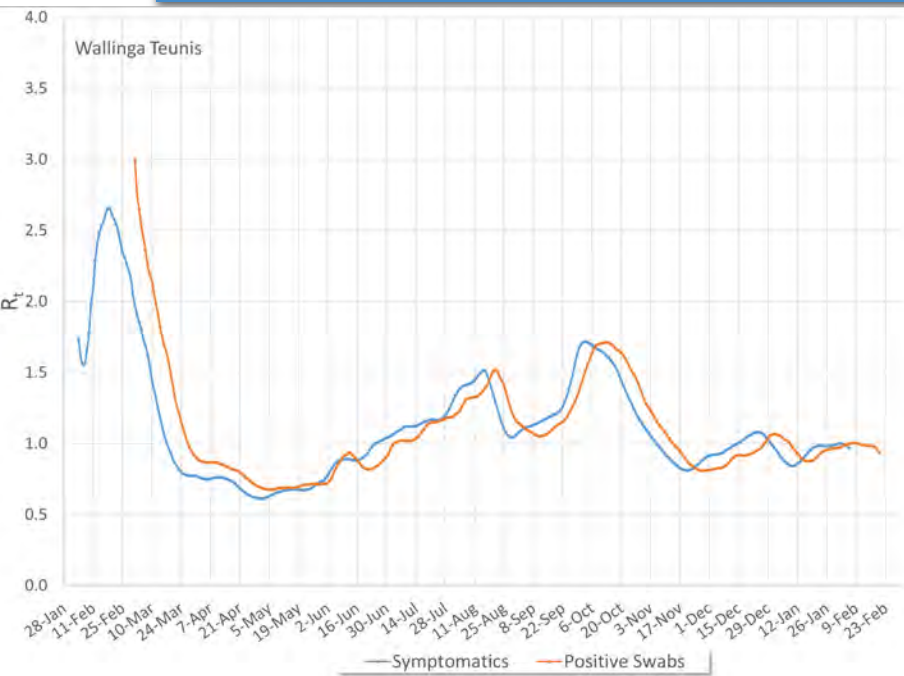
Gli errori sono relativi ai sintomatici e tengono conto degli errori statistici e della incertezza del generation time

R_t calcolato con i tamponi positivi è in perfetto accordo con R_t calcolato con i sintomatici

I sintomatici sono in anticipo di 8 giorni MA gli ultimi 14 giorni non sono utilizzabili.

Ne consegue che R_t con i tamponi positivi è in grado di produrre risultati con 6 giorni di anticipo

Confronto di R_t calcolato con altri 4 algoritmi



Qui non abbiamo anticipato i dati dei tamponi positivi

I tamponi positivi sono rilevati dal 25feb20. I dati dei sintomatici derivano dai tamponi positivi. Per questo motivo R_t dei sintomatici nei giorni precedenti al 25 febbraio 2020 non e' significativo.

E' comunque rilevante che dal 25feb20, prima del lockdown, R_t fosse gia' in discesa.

Cori et al. e' l'algoritmo utilizzato da ISS, per cui questo plot mostra come R_t dell'ISS verrebbe modificato (if any) utilizzando i dati dei tamponi positivi

Un altro dato interessante: frazione di asintomatici



Anticipando di 8 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.