COVIDSTAT INFN

INFN
Istituto Nazionale di Fisica Nucleare

Approfondimenti sulle tecniche di analisi utilizzate nel sito e i loro limiti

Curve per la descrizione dei dati integrali (fit globali)

L'analisi dei conteggi, una volta sommati al giorno precedente (**dati integrali**) , prevede numeri sempre crescenti fino alla fine dell'epidemia quando il numero finale non cresce più. Tipiche distribuzioni che descrivono questo andamento sono le **funzioni logistiche** oppure l'integrale di una gaussiana che prende il nome di **Error function (ERF)**

Es. di **funzione logistica**

$$N(t) = a\frac{1+me^{-t/\tau}}{1+ne^{-t/\tau}}$$

Usata ad es. per descrivere la crescita feto: crescita veloce iniziale poi limitata dalle dimensioni dell'utero

Ad ogni modello di crescita → equazione differenziale → soluzione

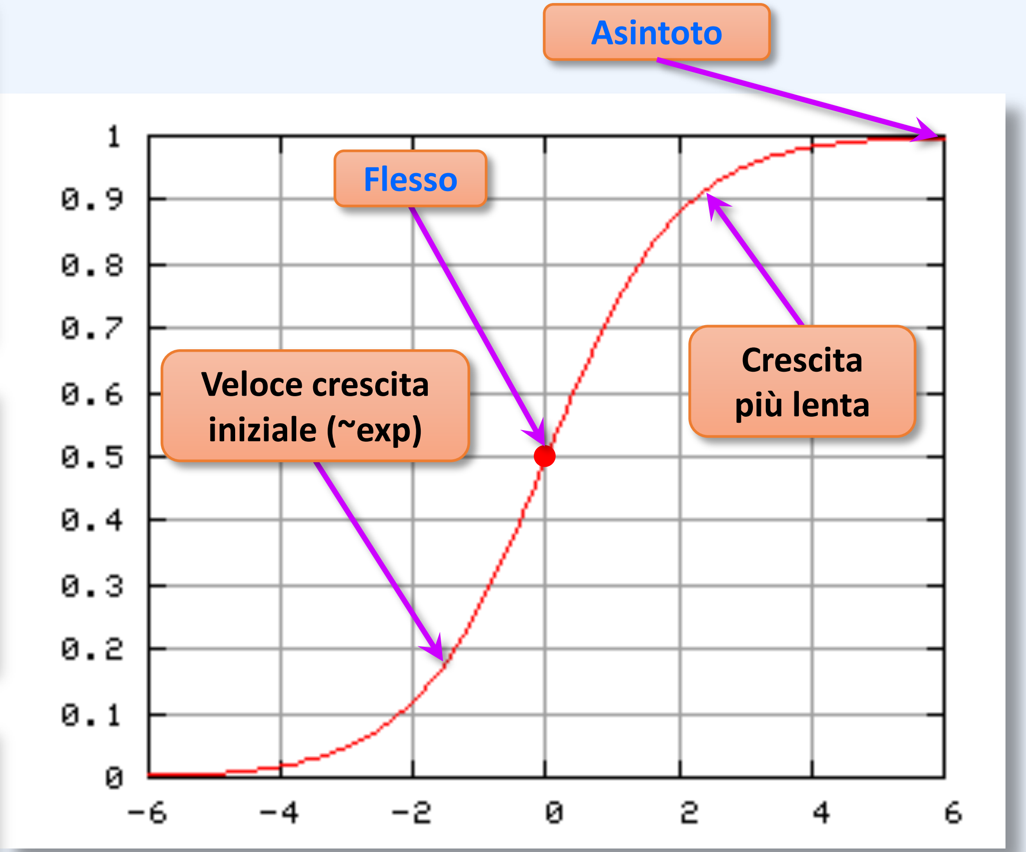
Particolare funzione logistica da noi scelta

$$\frac{dN}{dt} = rN\left(1-\frac{N}{K}\right) \qquad N(t) = \frac{K}{1+qe^{-rt}} \qquad q = \frac{K-N(0)}{N(0)}$$

N: popolazione
r: tasso di crescita
K: termine asintotico

funzione ERF (integrale della distribuzione normale detta Gaussiana)

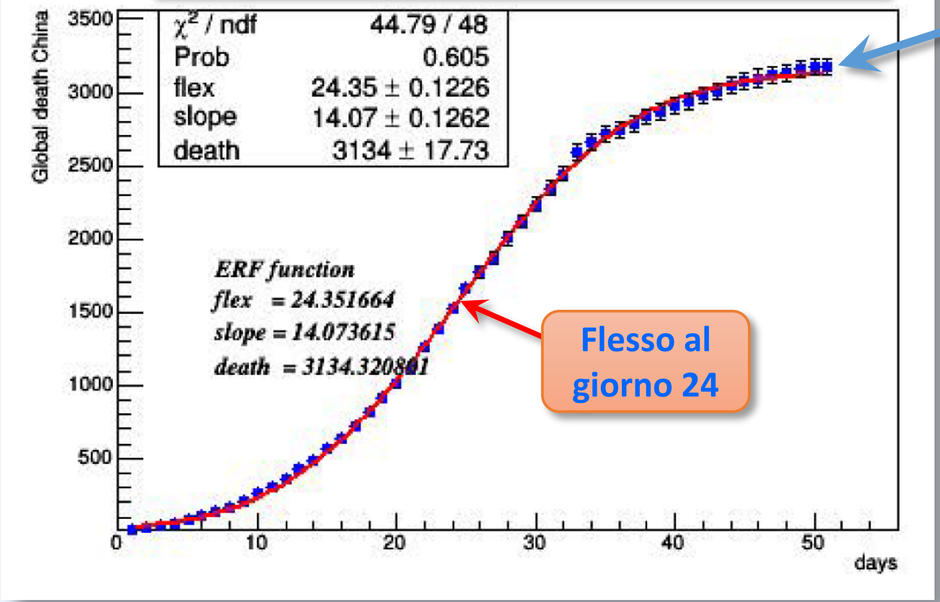
$$ERF = \int_{-\infty}^{+\infty} f(x)dx \qquad ERF\left(Asymp\frac{x-flex}{slope}\right)$$



Previsione dei fit globali durante l'epidemia

Istituto Nazionale di Fisica Nucleare

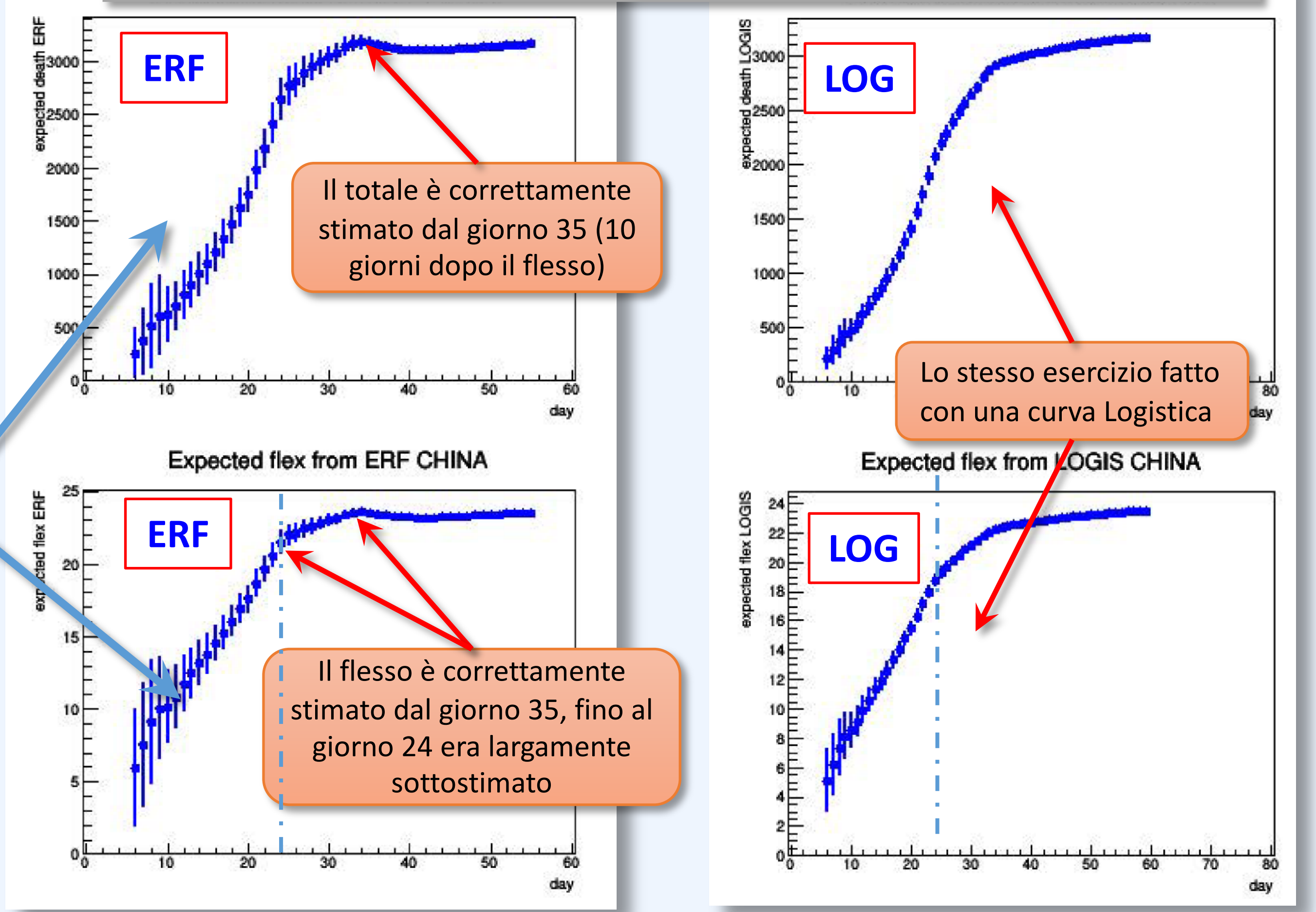
Numero totale dei deceduti in Cina



Flesso al giorno 24

Una curva ERF può interpolare a posteriori i dati della Cina sui decessi totali in modo eccellente ($\chi^2\ /ndof \cong\ 1$).

Domanda: può anche estrapolarli (ovvero prevedere la posizione del flesso e il totale prima che siano raggiunti)?



Il totale è correttamente stimato dal giorno 35 (10 giorni dopo il flesso)

Lo stesso esercizio fatto con una curva Logistica

Il flesso è correttamente stimato dal giorno 35, fino al giorno 24 era largamente sottostimato

Test cruciale: interpolazione al giorno "n" con una curva ERF, riportando per ogni giorno il valore stimato del flesso e del totale.

È evidente che la curva sottostima costantemente la posizione del flesso, e lo riproduce correttamente solo con tutti i dati a disposizione, quando il flesso è stato abbondantemente superato

Curve per la descrizione dei dati giornalieri (fit locali)

INFN
Istituto Nazionale di Fisica Nucleare

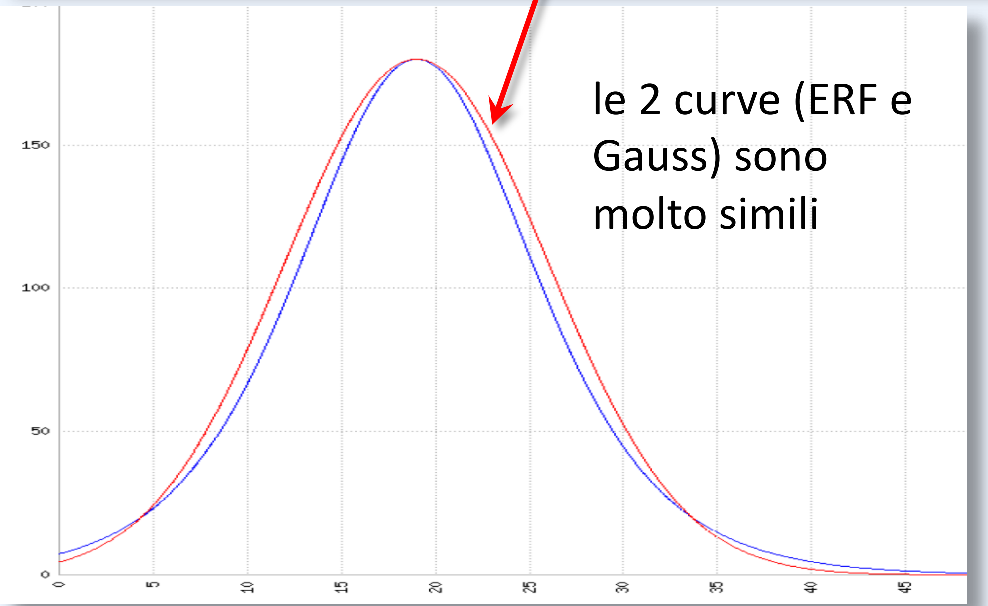
L'analisi dei conteggi giornalieri, prevede numeri che partono da zero, arrivano ad un massimo e poi tornano a zero alla fine dell'epidemia. Tipiche distribuzioni che descrivono questo andamento sono la **funzione gaussiana** e la **derivata delle funzioni logistiche**

Funzione gaussiana (Gaus)

$$f(x) = \frac{A}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2}$$

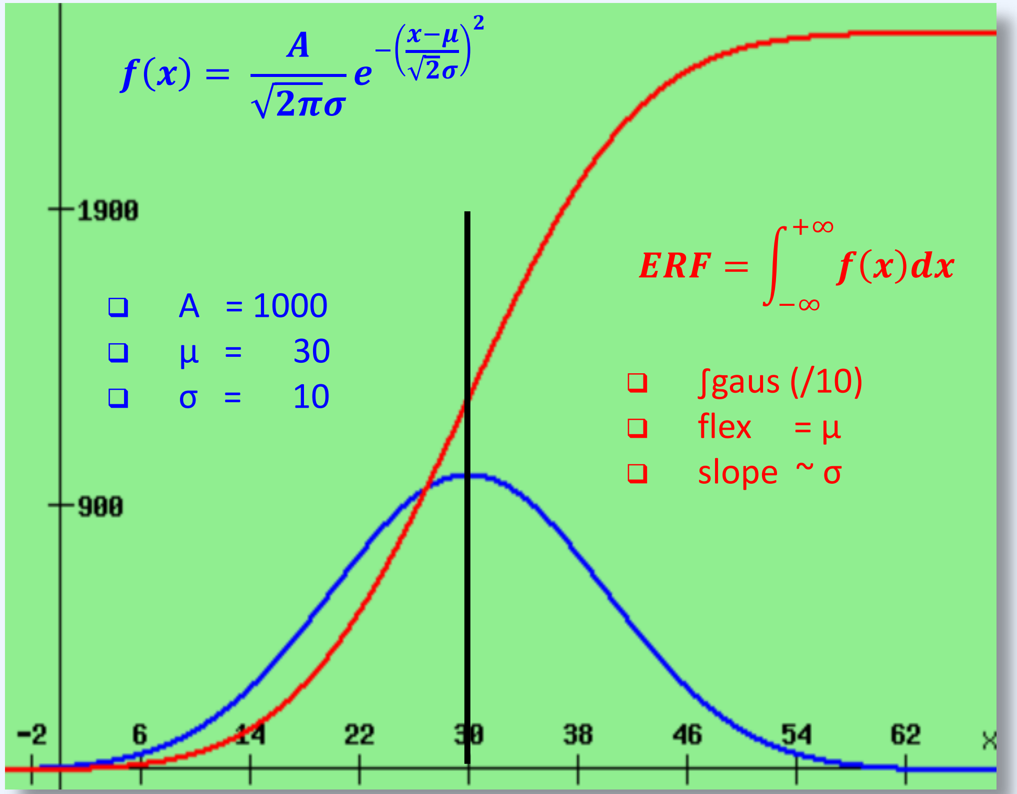
Derivata della funzione logistica (dLog)

$$\frac{dLogistica}{dt} = \frac{K * slope * e^{-slope*(t-flex)}}{(1 + e^{-slope*(t-flex)})^2}$$



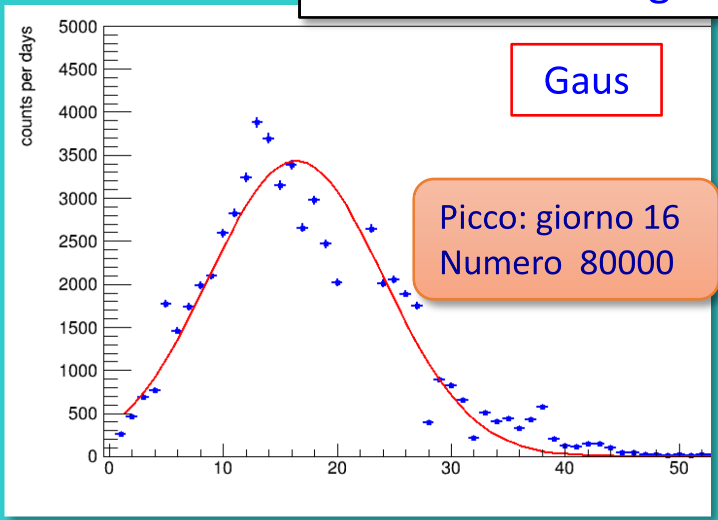
Relazione tra la funzione integrale ERF e la gaussiana

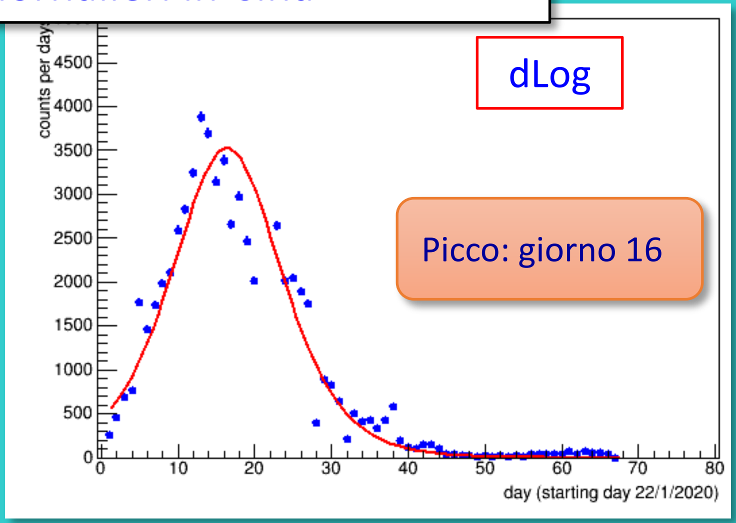
Picco (gaussiana) = **Flesso** (ERF)
Area (gaussiana) = **Asintoto** (ERF)



Stime dei fit locali a fine epidemia

Contagiati giornalieri in Cina

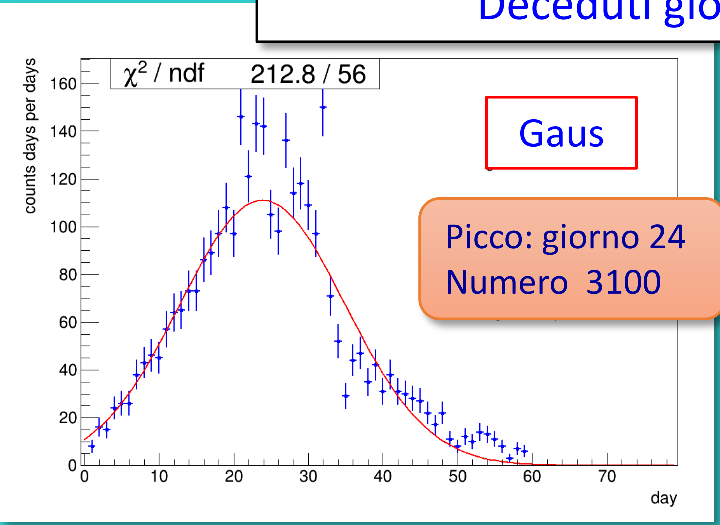




Numeri effettivi
- **Contagiati = 81054**
- **Picco = 16° giorno (6/2)**

Le 2 curve danno risultati compatibili

Deceduti giornalieri in Cina



Numeri effettivi
- **Decessi = 3261**
- **Picco = 24° giorno (14/2)**

Risultato:
- **Picco** (gaussiana) = **Flesso** (ERF)
- **Area** (gaussiana) = **Asintoto** (ERF)

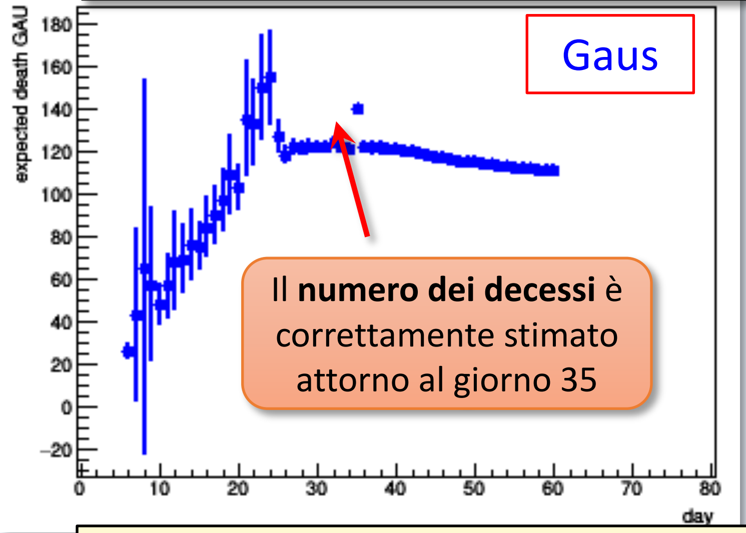
Stesso risultato con la curva logistica

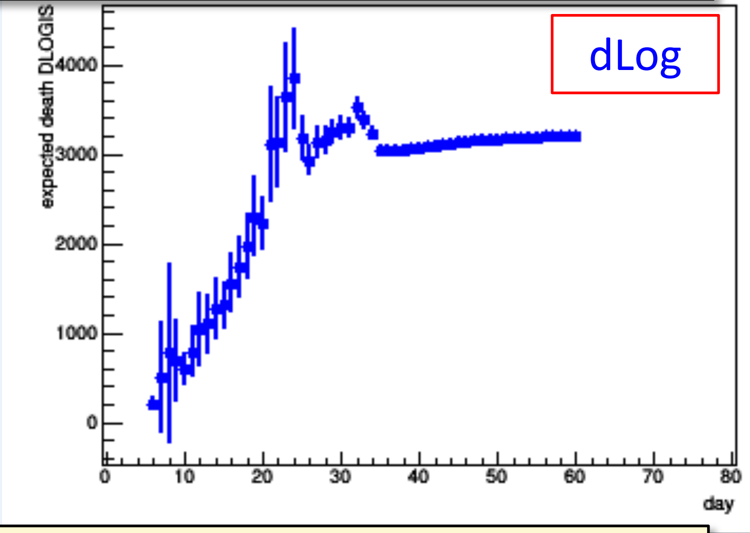
Fit Globali e locali ben descrivono l'epidemia cinese (60 giorni di misure)

Previsione dei fit locali durante l’epidemia

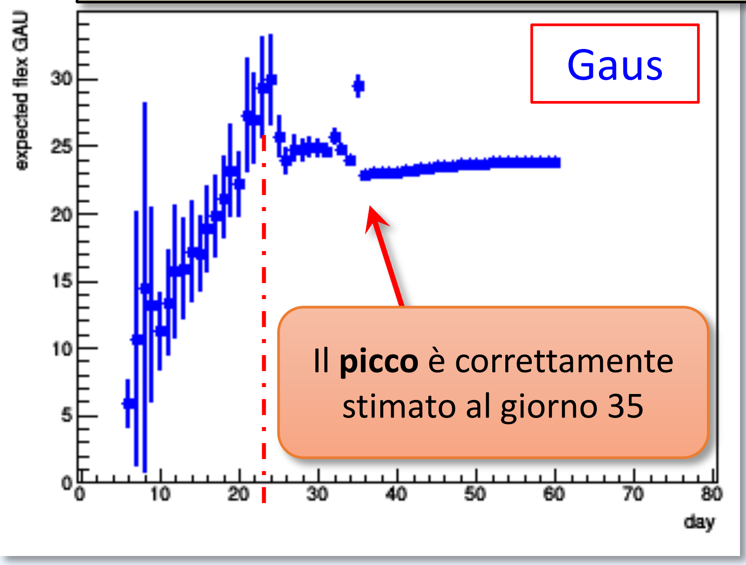
INFN
Istituto Nazionale di Fisica Nucleare

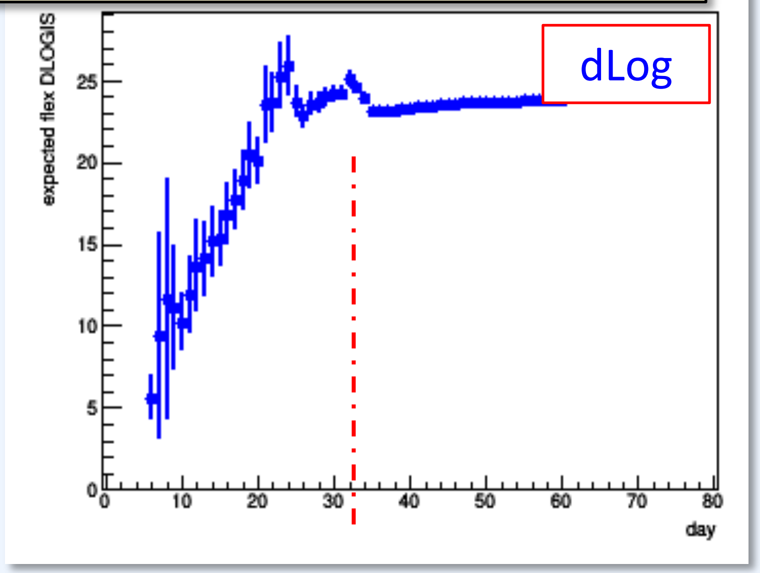
Numero atteso dei deceduti in Cina





Posizione del picco atteso per i decessi in Cina





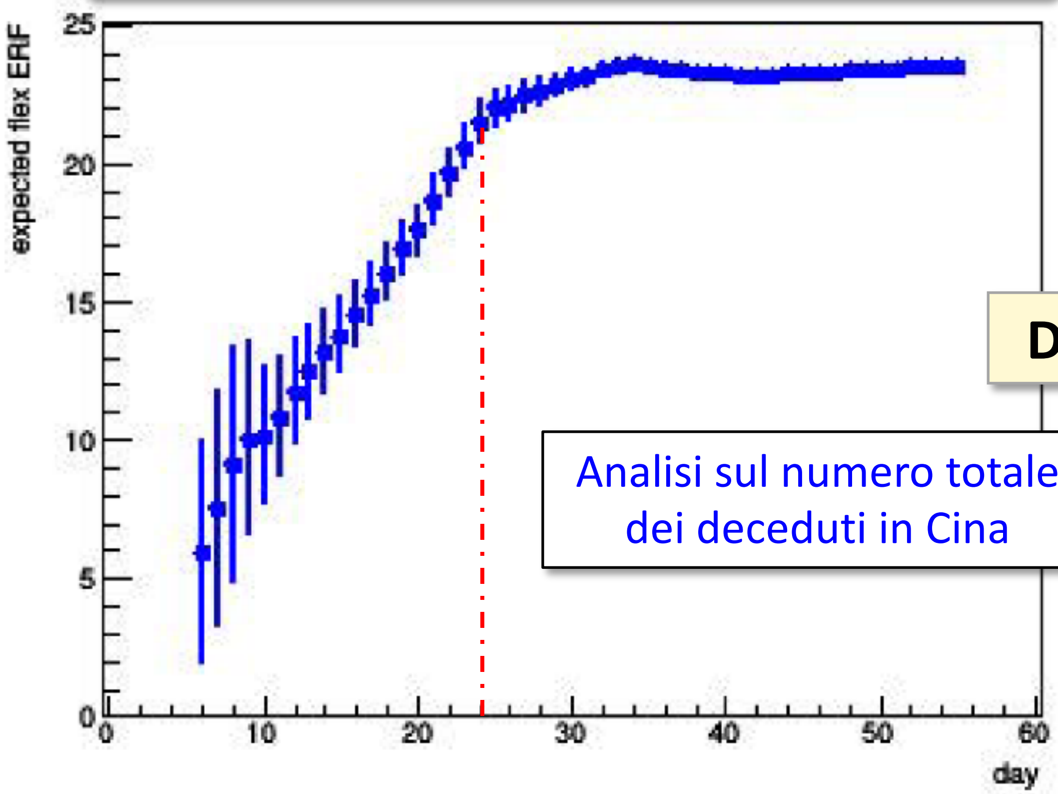
Per ogni giorno è stimata la previsione del fit sul numero dei decessi e sulla posizione del massimo

Il numero dei decessi e il picco sono correttamente stimati attorno al giorno 35 (10 giorni dopo averlo raggiunto)

Quando i dati avevano raggiunto il picco il fit non aveva ancora il corretto potere predittivo

Cina: confronto tra fit globali e locali

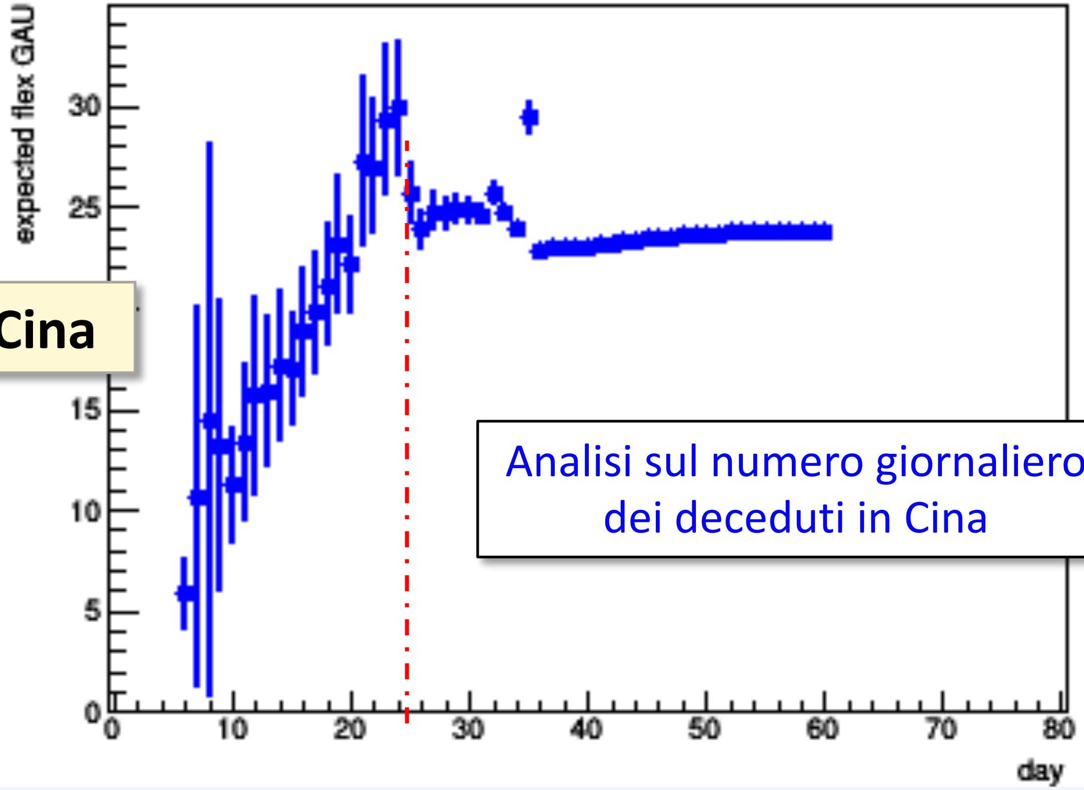
Previsione sulla posizione del flesso con la ERF



Analisi sul numero totale dei deceduti in Cina

Deceduti in Cina

Previsione sulla posizione del flesso con la GAUSSIANA

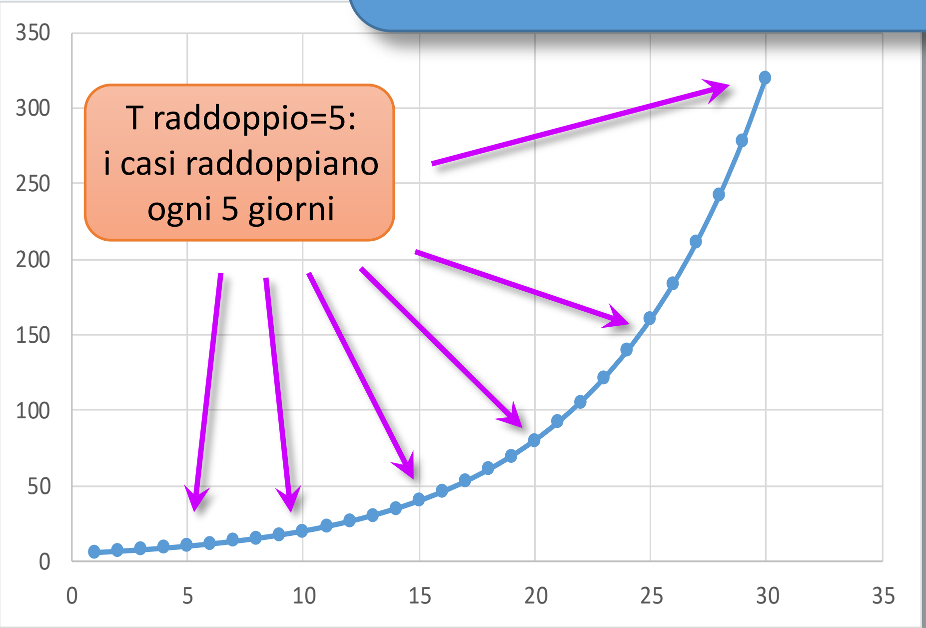


Analisi sul numero giornaliero dei deceduti in Cina

Dopo averlo oltrepassato, prevedono il corretto valore del flesso in modo diverso:

- **ERF** (cumulative) risente delle correlazioni: per qualche ragione il flesso era inizialmente sottostimato (e si ripercuote sui punti successivi) e viene monotonamente corretto nei giorni successivi
- **Gaussiana** (distribuzione giorno per giorno) risente solo delle fluttuazioni statistiche casuali, dopo il picco si avvicina al valore vero oscillandoci attorno

Curve esponenziali

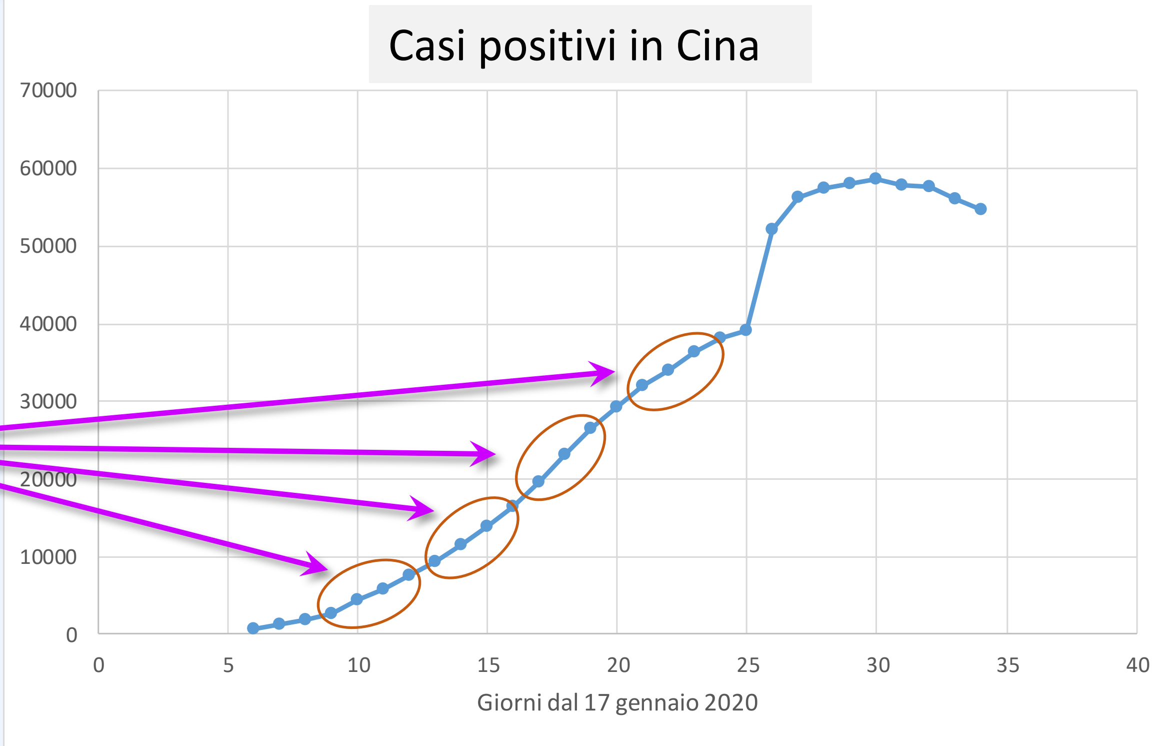


Una curva esponenziale y(t)=k·$e^{t/\tau}$ è descritta da due costanti: una normalizzazione «k» (che non è in questo caso interessante) e una costante di tempo «τ» inversamente proporzionale al tempo di raddoppio della curva: ad intervalli di tempo uguali i casi raddoppiano

Una crescita esponenziale è destinata comunque a saturare, non fosse altro perché ad un certo punto i casi positivi corrispondono a tutta la popolazione nazionale (con un tempo di raddoppio di 6 giorni, questa condizione sarebbe raggiunta in Italia verso la fine di Maggio)

Tutte le misure di contenimento si propongono quindi di passare da una crescita esponenziale ad una crescita più controllata, fino a raggiungere la fine dell'aumento dei contagi.

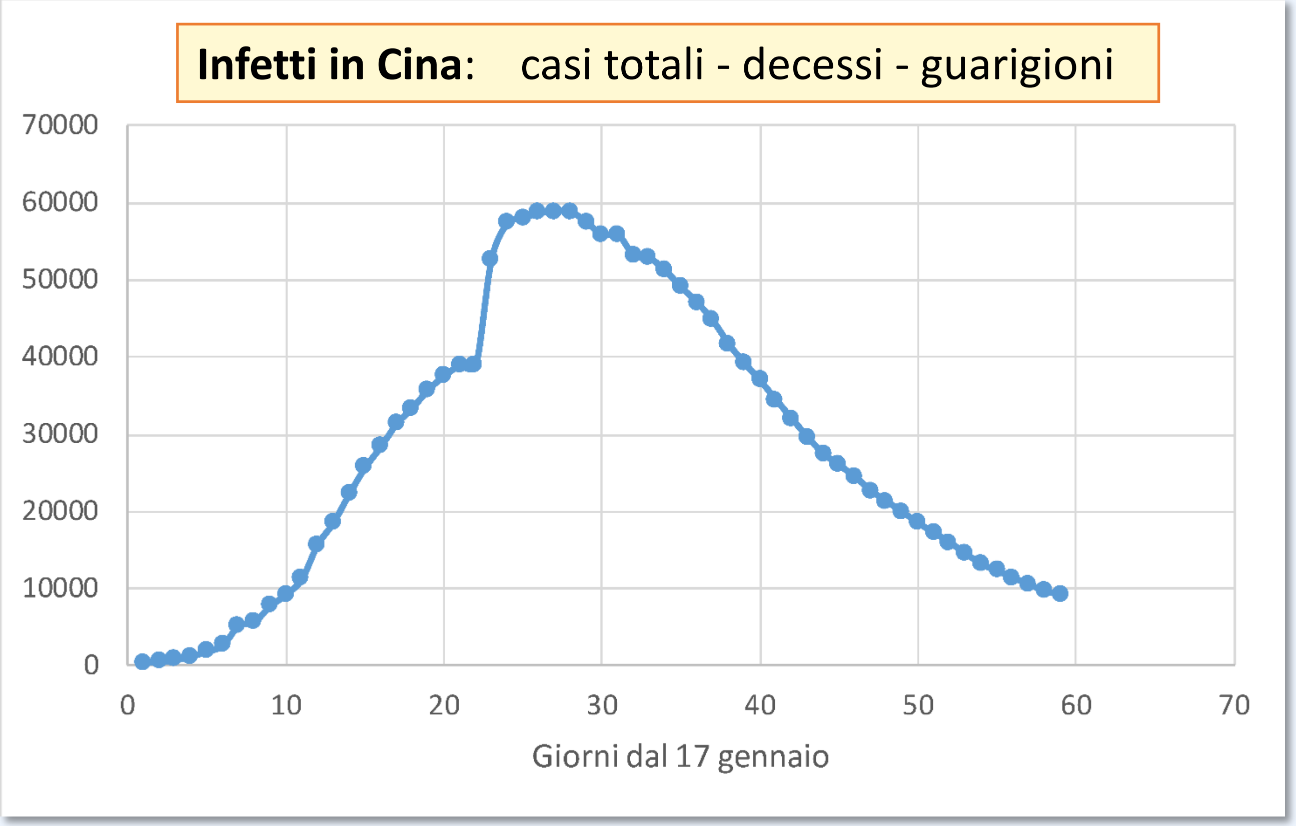
Fittando con una curva esponenziale solamente gli ultimi giorni a disposizione, è quindi possibile misurare l'effetto delle azioni di contenimento in termini di modifica (allungamento) del tempo di raddoppio.

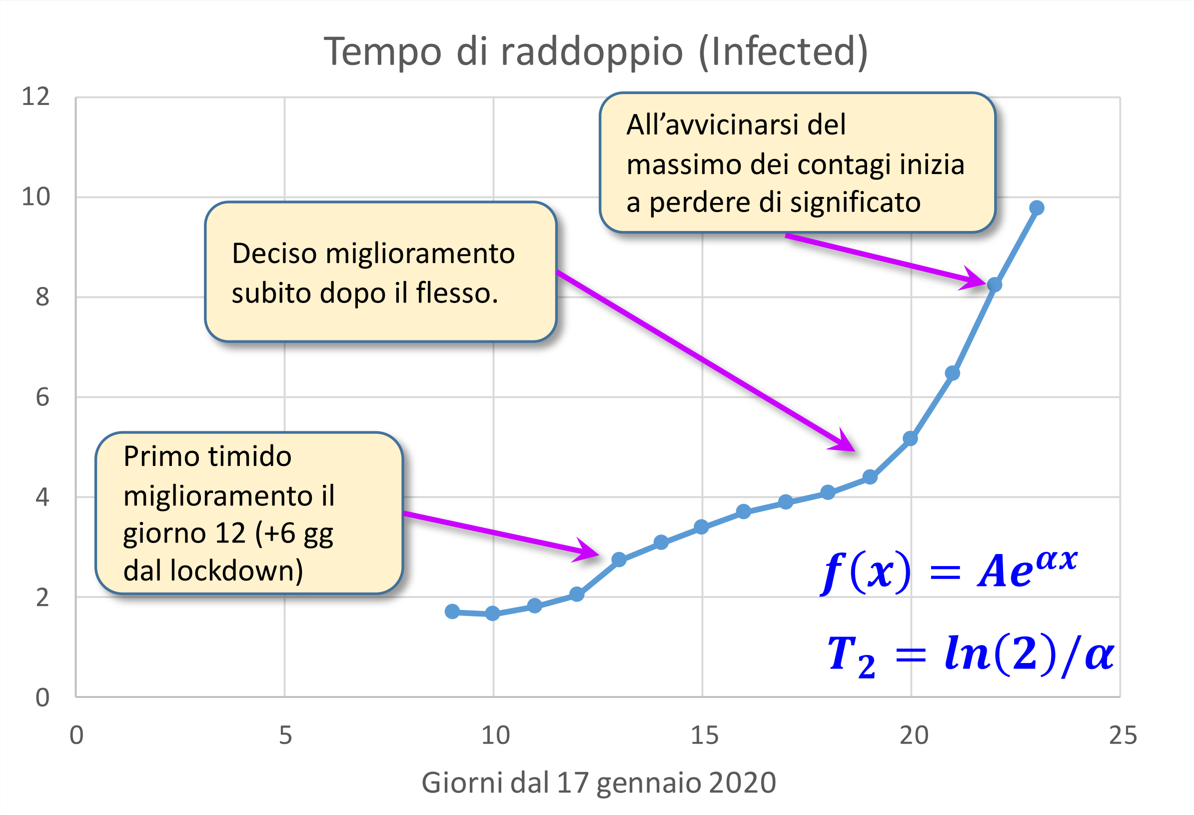


Gli effetti delle misure di contenimento diventano comunque visibili solo quando i casi positivi diventano misurabili, ovvero dopo il tempo necessario per il manifestarsi dei sintomi.

Tempo di raddoppio

- Interpolazione esponenziale mobile a 5 giorni su “***Infected***” o su “***Casi Totali***”. 5gg è un ragionevole compromesso fra l’assorbire errori statistici e sistematici e sensibilità ai cambiamenti.





A fine contagio la distribuzione tende ad appiattirsi e $T_2 \longrightarrow \infty$

Modelli SIRD, R_0 e R_{th}

Cercano di descrivere la dinamica di

- **S**(usceptible): popolazione suscettibile al contagio
- **I**(nfected): contagiati tuttora infetti, con rate β(I)
- **R**(ecovered): guarigioni, con rate α(I)
- **D**(ead): decessi, con rate γ(I)

$$R_{th} = \frac{\beta}{\alpha + \gamma}$$

R_{th} è un possibile stimatore di R_0 che ha un preciso significato epidemiologico: rappresenta il numero medio di contaminati da un singolo soggetto infettivo. Un valore di $R_0 \leq 1$ significa che l'epidemia è ormai contenuta

Il modello semplificato da noi utilizzato e i dati a nostra disposizione non ci permettono di stimare correttamente R_0. Abbiamo comunque messo alla prova R_{th} con i dati cinesi per valutare le sue capacità di descrivere il contagio.

R_{th} viene calcolato assumendo che

- $S = \infty$
- Le rate vengono mediate su 4 giorni
- Non viene applicate alcuna correzione sul fatto che le variabili in gioco si sviluppano con ritardi diversi