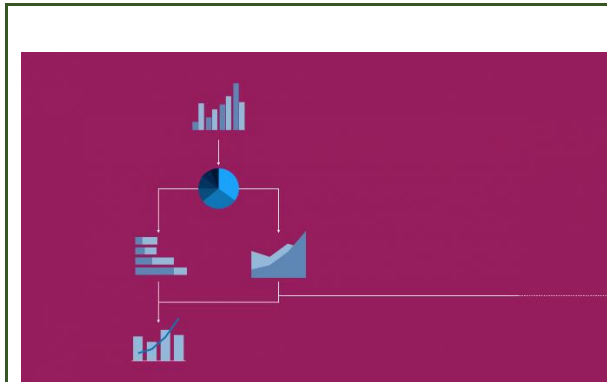


## 96. Scienza&Società L'insostenibile delicatezza dei modelli previsionali\* Covid-19



I modelli previsionali usati per capire l'andamento di Covid-19 sono soggetti a un margine d'incertezza ed è molto difficile prevedere, con livello accettabile di affidabilità o limiti stretti di incertezza (che andrebbero sempre riportati), quando inizierà la riduzione significativa della velocità di crescita e quando si raggiungerà l'apice e la discesa. Questo li rende materia di comunicazione estremamente delicata perché, oltre alle distorsioni del significato, si possono ingenerare reazioni impattanti sulle misure di sanità pubblica.

I modelli previsionali usati per capire l'andamento di Covid-19 sono soggetti a un margine d'incertezza, ed è molto difficile prevedere, con livello accettabile di affidabilità o limiti stretti di incertezza (che andrebbero sempre riportati) quando inizierà la riduzione significativa della velocità di crescita e quando si raggiungerà l'apice e la discesa. Questo li rende materia di comunicazione estremamente delicata perché, oltre alle distorsioni del significato si possono ingenerare reazioni impattanti sulle misure di sanità pubblica.

Sono talmente d'accordo con chi, [anche su Scienza in rete](#), ha fatto presente le incertezze e la delicatezza di uso in questa fase dei modelli previsionali, che non intendo iscrivermi tra coloro che pure premendo le incertezze poi non riescono a rifuggire la tentazione di dare qualche previsione.

Come non-modellista ma utente dei risultati dei modelli sono preoccupato della apertura di una fase in cui vengono "buttate" sul terreno analisi e risultati differenti, non tanto per la diversità di risultato, ma soprattutto per la mancanza di chiarezza dello scopo, e da qui delle altamente probabili distorsioni sul piano comunicativo.

### Margini d'incertezza

Fare previsioni sul breve termine non è difficile: per ogni distribuzione temporale si può trovare la migliore curva che interpola l'andamento a oggi (modelli empirici dinamici), e da qui prevedere il domani, tanto più agilmente quanto più l'andamento è semplice. L'esercizio di offrire curve di andamento attraverso l'uso di modelli matematici empirici, ad esempio le funzione esponenziali [offerte da Alessandro Marinaccio](#) e quelle logistiche o più complesse [da Giovanni Sebastiani](#) su *Scienza in rete*, penso vada iscritto in questo ambito, con la consapevolezza da parte degli autori di non dare informazioni di lungo respiro, se non accettando **ampi margini di incertezza**, in quanto trattandosi di modelli descrittivi i parametri (in primis il tasso di crescita) hanno significato solo nelle condizioni sperimentali, nel nostro caso osservazionali, in cui sono stati determinati.

Con questi modelli penso sia ragionevole non spingersi oltre la segnalazione di una modifica del tasso di crescita, come del resto fanno gli autori. Infatti le assunzioni non mancano, anche forti: che i guariti non contengano più il virus, o che comunque non possano trasmetterlo ad altri individui infettandoli, che non tutti gli individui diagnosticati infetti siano isolati, a causa di mancata o erronea diagnosi e/o isolamento, che

ci sia una compensazione degli errori che affliggono le misure delle singole variabili, esempio il numero dei morti, oltre alle carenze di dati come una stima attendibile della frazione dei portatori sani.

Molto più difficile è prevedere, con livello accettabile di affidabilità o limiti stretti di incertezza (che andrebbero sempre riportati) **quando inizierà la riduzione significativa della velocità di crescita** e quando si raggiungerà l'apice e la discesa, che sono le domande chiave di cittadini e amministratori. Qui l'impresa si fa più ardua, non per carenza di modelli super-sofisticati e per la disponibilità di modellisti di valore, ma per la necessità di alimentare questi modelli con dati "buoni", non solo quelli rilevati fino a oggi, ma anche quelli storici riferiti a pandemie ed epidemie del passato relative a malattie dello stesso tipo.

Siccome abbiamo a che fare con un nuovo virus dobbiamo essere consapevoli che i modelli comunemente utilizzati sono stati **tarati su pandemie vecchie** e che, d'altra parte, i dati dell'epidemia in Cina offrono la possibilità di tarare modelli pertinenti al Covid-19. Dal primo di marzo ho letto diversi articoli sulle previsioni su Covid-19 in Cina e alcune previsioni inviatemi da colleghi matematici, igienisti, epidemiologi e gruppi misti, alcuni su intervalli di tempo già trascorsi e per i quali è possibile verificare gli scostamenti delle previsioni quando non i fallimenti. I modelli utilizzati erano di tutti i tipi, meccanicistici, fenomenologici, con curve esponenziali, logistiche, polinomiali, sigmoidali, miste.

D'altra parte, la volontà e insieme il "potere" di prevedere il futuro ha accompagnato la storia dell'umanità, passando dalla "predizione", consistente nell'annunciare e fare profezie con l'aiuto di divinità, alla "previsione", cioè fare ipotesi e supposizioni a partire dai dati disponibili. Oggi, la disponibilità di una grande mole di dati, di modelli matematici raffinati e di potenti strumenti di calcolo permette a molti gruppi di ricerca e moltissimi ricercatori singoli di cimentarsi per produrre previsioni.

### **Sfidare le previsioni**

Oltre alle funzioni istituzionali che molti gruppi devono assolvere, c'è da parte del ricercatore un impulso a usare le proprie competenze e il proprio "ingegno" per contribuire alla crescita delle conoscenze, a mettere alla prova l'esercizio del dubbio, a sfidare il proprio "potere", e non c'è campo di applicazione più sfidante di quello delle previsioni. Ritengo paradigmatico uno [studio](#) previsionale su dati cinesi, suddivisi tra la provincia di Hubei e il resto della Cina, effettuato da ricercatori statunitensi e canadesi, con tre diversi modelli fenomenologici (crescita logistica generalizzata, crescita di Richards e onda sub-epidemic) usati precedentemente per fare previsioni su SARS, Ebola, influenza pandemica e dengue.

Inserendo i dati ufficiali dei contagiati confermati (fonte Chinese National Health Commission) dal 22 gennaio al 4 febbraio e calibrando i modelli con i dati osservati dal 5 al 9 febbraio, sono stati stimati i numeri cumulativi di casi a 5, 10 e 15 giorni, cioè al 14, 19 e 24 febbraio. I modelli davano risultati abbastanza concordanti tra loro ma con tre stime del 34%, 40% e 41% in meno rispetto ai dati che si sono poi realmente verificati.

E ci si chiede subito quale possa essere stata la causa di sottostime tra 17.000 e 27.000 casi positivi a Covid-19. Si può ben dire che i ricercatori non sono stati fortunati, perché nei giorni successivi alla loro calibrazione, il 13 febbraio, le autorità cinesi hanno deciso di aggiungere ai casi confermati su tampone quelli da diagnosi clinica, tanto è che il 14 febbraio hanno avuto un incremento giornaliero del 50% a fronte del 4-5% dei giorni precedenti. La stima rispetto ai soli dati confermati avrebbe fallito del 4%, mentre per i giorni successivi i dati comunicati all'OMS non erano più disaggregati secondo conferma o sospetto e non è più possibile calcolare esattamente l'entità della sottostima. Una stima basata sul fatto che i sospettati clinicamente erano stati fino ad allora intorno al 30% del totale permette di predire circa il 10% di sottostima nei giorni 19 e 24 febbraio.

Morale della favola, si può constatare la conseguenza negativa del cambiamento di protocollo di rilevazione/registrazione a "treno in corsa" e la estrema fragilità dei modelli a cui non si può certo chiedere di tener conto dei cambiamenti apportati nel sistema di sorveglianza.

Un ragionamento analogo mi pare valga anche per le differenze geografiche e spazio-temporali (situazioni geografiche sfalsate nel tempo), anche di protocolli di definizione e registrazione delle diverse tipologie di caso, che indicano la inadeguatezza di confronti e anche di uso di modellistiche previsionali su ambiti eterogenei al proprio interno. Questo non riguarda solo differenze tra nazioni ma anche tra regioni, come mostrato anche dalle [esercitazioni di Giovanni Sebastiani](#) su *Scienza in rete*. L'eterogeneità spazio-

temporale del sistema di sorveglianza pandemica sarà sicuramente – o auspicabilmente - un argomento di attenzione e azione da parte dell'OMS.

### **Multidisciplinarietà e delicatezza della comunicazione**

Per concludere, ancora due considerazioni sui modelli e il loro uso in sanità pubblica. La prima è che i modelli matematici hanno bisogno di **gruppi con competenze multidisciplinari** in cui il matematico o il fisico, lavorano a stretto contatto con epidemiologi, infettivologi, clinici operatori e manager di sanità pubblica, avendo al proprio interno chi conosce esattamente limiti e potenzialità dei dati prodotti (chi produce i dati insieme a chi li analizza).

La seconda è che la comunicazione dei risultati dei modelli, una volta ridotta l'incertezza della stima, è **estremamente delicata**, perché, oltre alle distorsioni del significato si possono ingenerare reazioni impattanti sulle misure sociali e di sanità pubblica (contenimento). Si pensi ad esempio alla gestione comunicativa di un eventuale allontanamento della fase di decrescita dopo che è stato dato un periodo cui milioni di persone si "affezionano" (esempio metà aprile, avanzato da diverse parti) dell'arrivo dell'apice della curva epidemica, che tutti attendono con la speranza di tornare a una vita "normale", di possibili fenomeni di reinfezione e di molti altri punti critici a seguire.

Penso che sul piano comunicativo si possa ragionevolmente ipotizzare che una spiegazione in termini divulgativi non solo dei dati ma anche di criteri e meccanismi con cui si formano (importanza di metadati e metodi) potrebbe svelare il perché vengono prese certe decisioni, togliendole da accuse o sospetti di aleatorietà e accrescendo il rapporto di fiducia tra cittadini e istituzioni. Una misura di cui ci sarà molto bisogno per passare questa fase e non meno in quella successiva.

Oggi non è che un giorno qualunque di tutti i giorni che verranno, ma ciò che farai in tutti i giorni che verranno dipende da quello che farai oggi. È stato così tante volte.

*Per chi suona la campana, Ernest Hemingway*

\*Pubblicato in [scienzainrete](http://scienzainrete.it), 16 marzo 2020

18 marzo 2020  
Codice ISSN 2420-8442