On the uncertainty of real-time predictions of epidemic growths: A COVID-19 case study for China and Italy

ECS | Highlight

Tommaso Alberti[1,*] and Davide Faranda[2]

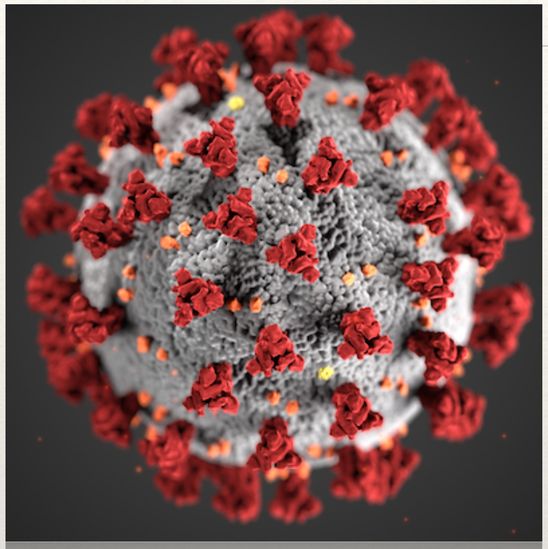
[1] INAF - Institute for Space Astrophysics and Planetology, Rome, Italy

[2] Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, IPSL and Université Paris-Saclay, Paris, France

* tommaso.alberti@inaf.it

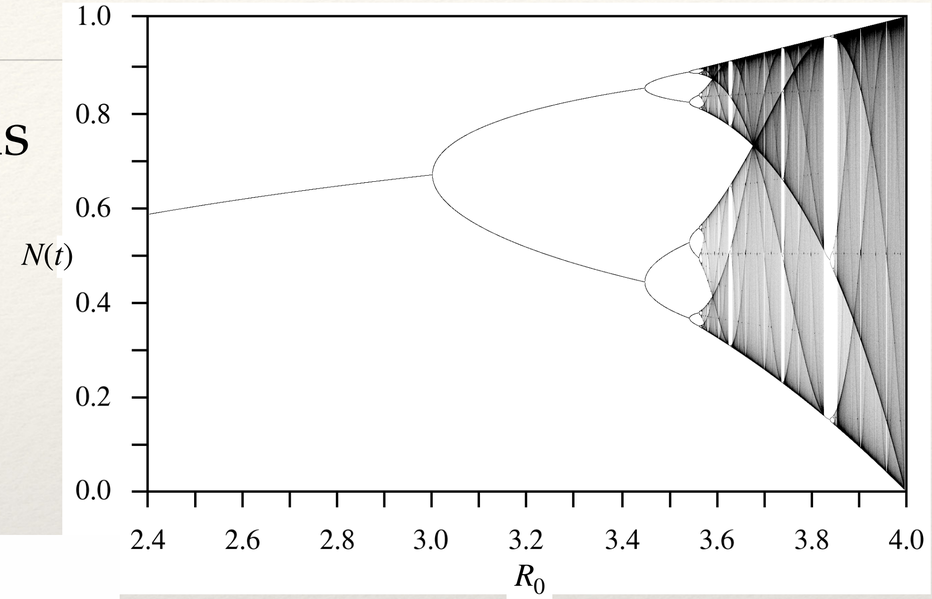
EGU General Assembly 2021

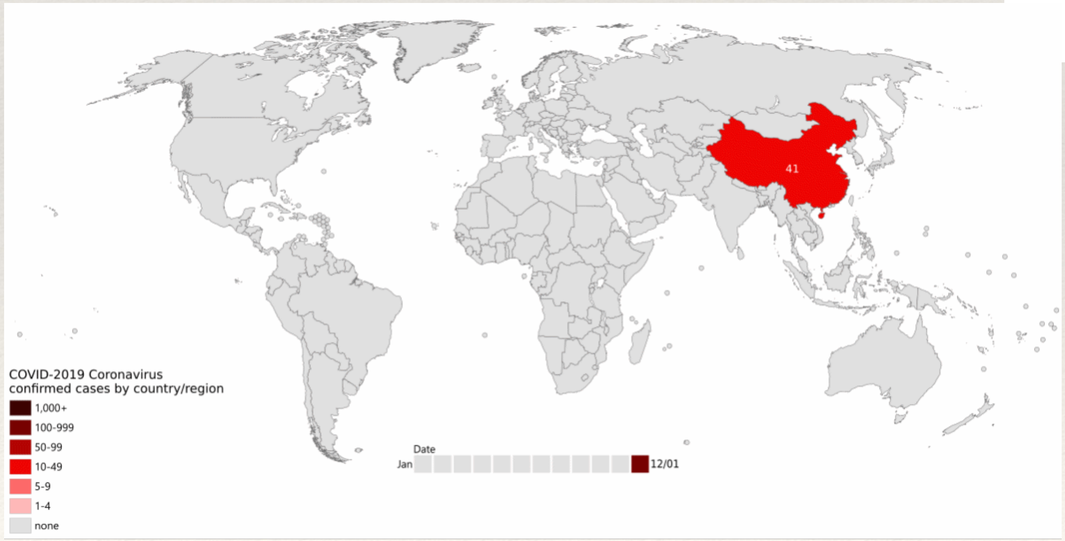
COVID-19: a complex system



COVID-19 pandemic evolution is a multiscale complex system

$$\frac{dN(t)}{dt} = R_0\, N(t) \left(1 - \frac{N(t)}{N_\infty}\right)$$





Network:
move one infected person from a region to another

Nonlinear:
take care of R_0

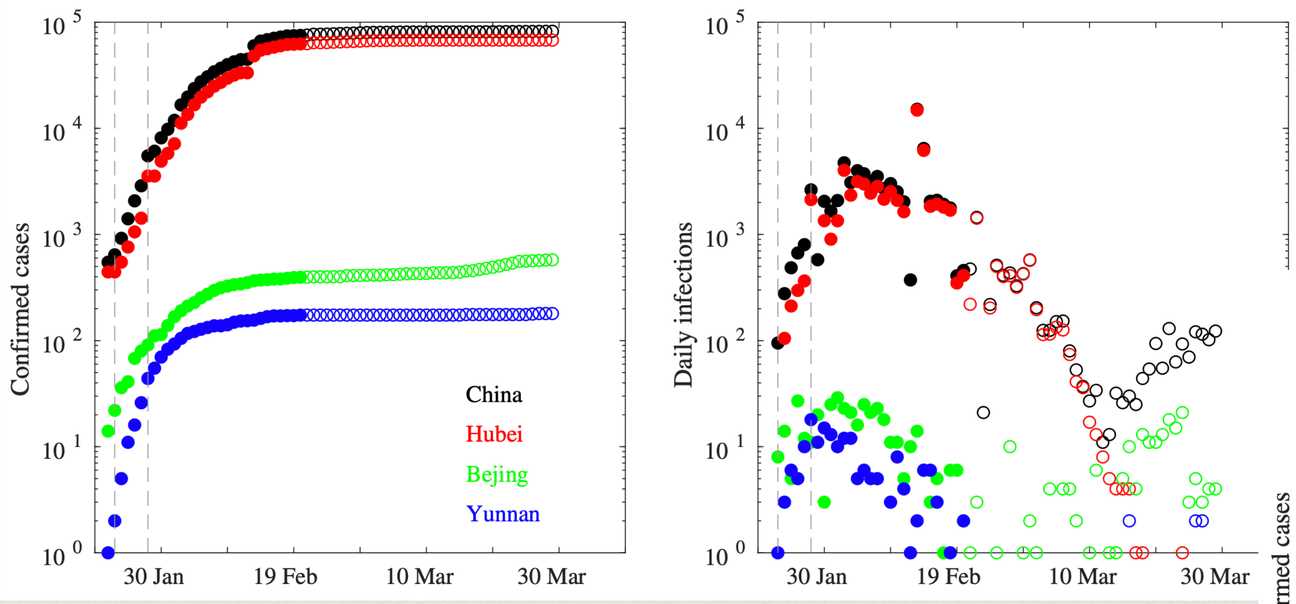
Self-organization:
Take order from disorder via local interactions

unfortunately here order means "infect all people"

Alberti and Faranda, Comm. Nonlin. Sci. Num. Sim., 90, 105372, 2020

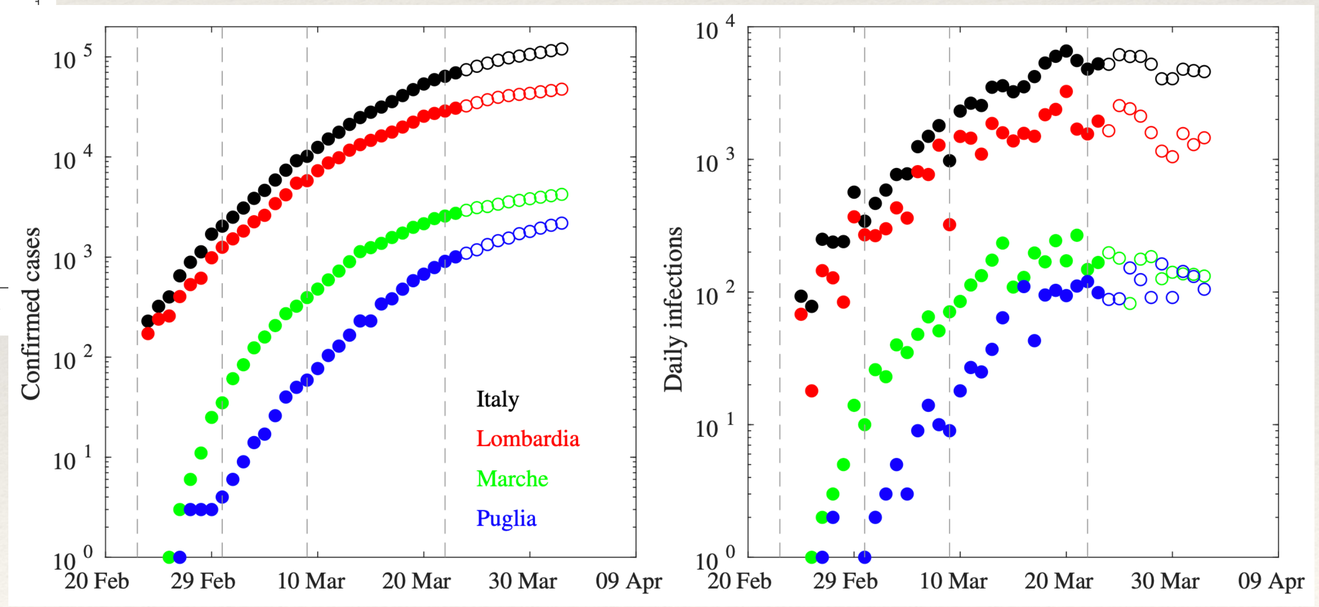
The first wave in China and Italy

- We used data collected during the period between 22 January and 30 March



China and some selected Chinese provinces

Italy and some selected Italian regions



Generalized logistic distribution

We perform logistic fits of the form

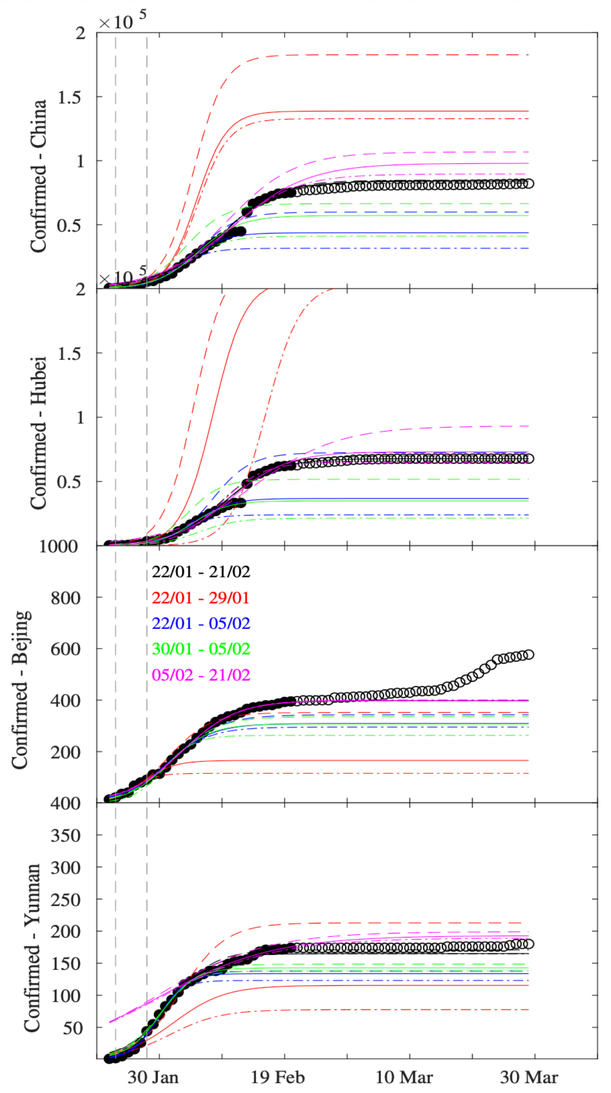
$$C(t) = \frac{\alpha}{1 + \beta e^{-\gamma t}}$$

by selecting the following time intervals:

- the first 30 days of epidemic growth, thus to consider how restrictions measure globally affect the diffusion;
- the first 7 days, roughly corresponding to the time interval during which first restriction measures are adopted both in China and Italy, although not still completely efficient;
- the first 14 days, corresponding to the time interval in which the initial confinement measures should lead the first effects;
- the time interval between the 8th and the 14th day to investigate how the epidemic would be grown if starting from initial restrictions;
- the time interval between the 15th and the 30th day to investigate the efficiency of restriction measures.

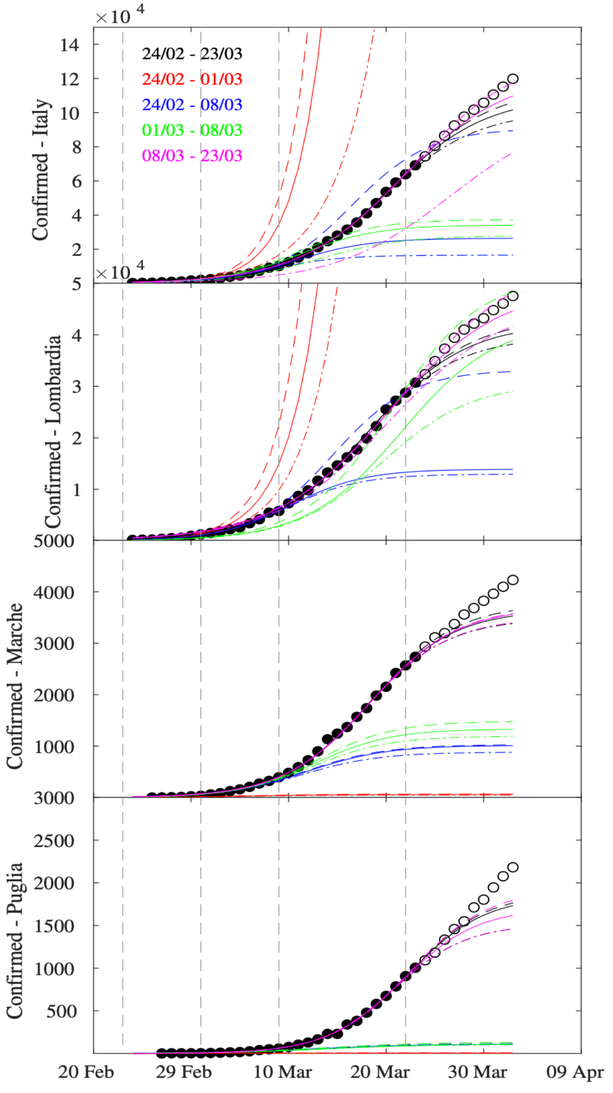
In this way we can investigate both the efficiency of restriction measures in containing epidemic growth as well as the stability of prediction models.

Logistic fits for China



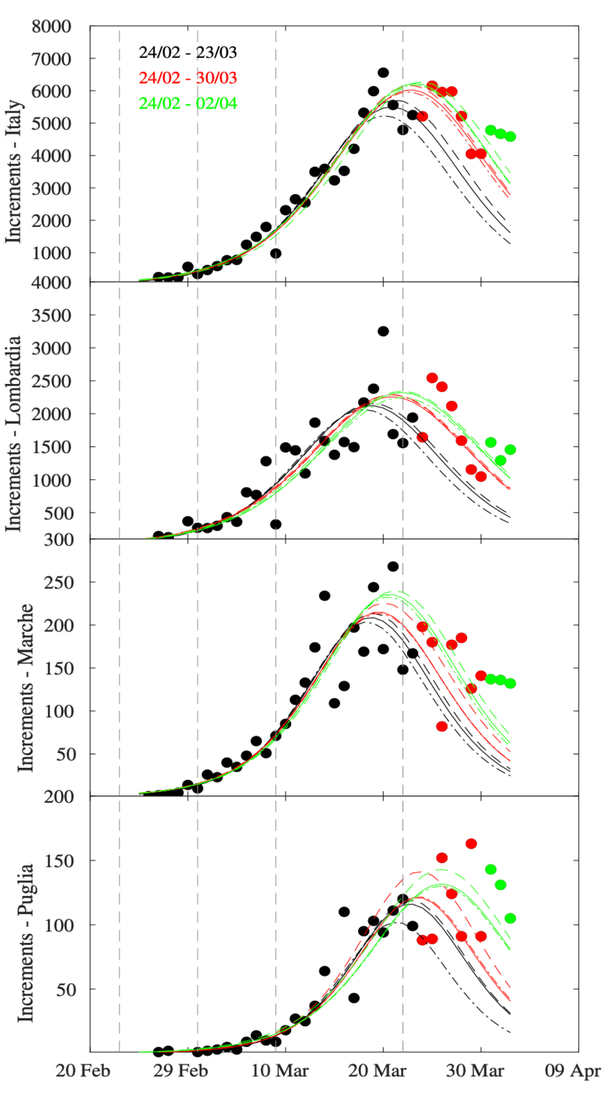
- early stage of epidemic propagation is characterized by a larger confidence interval (red lines), thus highlighting the difficulty in making early reliable predictions of epidemic growth, with an exponential-like behavior
- logistic fit becomes more stable, being characterized by a narrower estimates of confidence intervals, when the first two weeks are considered (blue lines), possibly related to the initial efficiency of restriction measures
- stability significantly increases if we do not include the first week of the epidemic growth (green and magenta lines), suggesting that credible predictions could be assigned with a large confidence by means of a logistic fit if the beginning of the outbreak is not considered

Logistic fits for Italy



- again reliable predictions of epidemic growth are particularly difficult in its early stages (red lines, mostly exponential)
- confidence intervals become narrower as the growth rate reduces (see for example Marche or Puglia with respect to Lombardia), with the logistic fits also becoming more stable when the initial stages of the outbreak are removed (green and magenta lines)
- Unlike for Chinese provinces, Italian regions present a wide range of different epidemic behaviors, due to the time delayed propagation of epidemic throughout Central and Southern regions for which a mature stage was not yet reached on March 30
- This result was also statistically assessed by using the Kolmogorov-Smirnov (K-S) test

Can we make reliable predictions?



- We tried to use logistic fits between 24 Feb-23 Mar (black line) to predict the behavior for the successive week 23 Mar-30 Mar (red line) as well as also including the next 3 days (30 Mar-02 Apr, green line)
- Lombardia presents a narrower confidence interval when including the successive days, not observed for both Marche and Puglia
- Particularly for Puglia the confidence interval remains practically unchanged, thus suggesting that logistic fits are not still stable
- This is possibly due to the fact that Southern regions have not yet reached a mature stage of the epidemic growth (and this was what happened!)

Tips & Conclusions

- By comparing our estimates and data collected from the daily report of the Italian Protezione Civile (https://github.com/pcm-dpc/COVID-19) we found that the discrepancy significantly increases when moving from Northern to Southern regions
- For Southern regions the error is comparable with the predicted value. This could be the reflection of at least two different factors:
    1. the epidemic growth is in a more mature phase in the Northern and Central regions with respect to the Southern ones, where it began with a time delay ranging from 3 to 14 days, and
    2. the higher ratio between the observed cases and the number of tests carried out for Southern regions with respect to the rest of Italy (see https://github.com/pcm-dpc/COVID-19).
- Thus, our results suggest that estimates of the ending of epidemic growth are affected by the statistical uncertainties, by the delayed propagation of infections through the different regions, and by the effective respect of the guidelines in terms of confinement measure!
- We highlight that reliable estimates cannot be released until more mature stages of the epidemic growth are reached.
- As a guideline it is surely helpful to perform logistic fits every day and to evaluate the reliability on predicting the next day, and then perform a new logistic fit to investigate how the uncertainty grown/reduced.
- The statistical modeling of epidemic growth should be focused on specific stages of its evolution on time as well as on its spread at a more local level!