

Supplementary Material

5 **This PDF file includes:**

Supplementary Material 1-Materials and Methods

Supplementary Material 2-Data correction for country inter-comparison

Supplementary Material 3-Dynamical regime

Supplementary Material 4-Case fatality rate

10 Supplementary Material 5-Canonical models

Supplementary Material 6-Italy Covid-19 model

Supplementary Material 7-Comparative analysis

Figs. S1 to S6

Tables S1 to S4

15

20

Supplementary Material 1-Materials and Methods

1. Context

As for numerous infectious diseases, several of the variables involved in the dynamic of the Covid-19 epidemic, although possibly obvious, are known to be quite difficult to observe: for diseases such as Covid-19, infectious people may not develop any clear symptoms, and susceptible people may be detected only once infected by the disease. Another compartment difficult to observe but commonly met in epidemic and surely involved in the Covid-19 epidemic is the compartment corresponding to exposed stage which relies to people having been infected but whom are not yet infectious themselves. Population may also be separated in classes depending on their age or social status to account for specific susceptibility due to it, or to model the transfer between different groups. It was found in particular for the Covid-19 disease that elder people were more susceptible to develop severe stages of the disease whereas severe stages were relatively rare among children whom may often be healthy carriers; It was also observed that health-care personal could constitute a specific class of people with higher susceptibility due to their direct exposure to the infectious people. The use of external compartments may also be useful to express the influence of social events that may strongly contribute to the epidemic, such as the travels associated to the New Year in China or sports and religious gatherings in France or Italy. Couplings to wild or domestic species may also play an important role, though, for the present epidemic, this coupling may probably have taken place at the earlier development of the epidemic only, and is no longer in progress. In some contexts, the anthropological dimension may also have an important influence on the epidemic (as it could be observed for other diseases such as Ebola Virus Disease), although few elements could be established clearly yet in this sense. The situation of the health systems may also play an important role in the monitoring of

the epidemic, both individually in its ability to deal with complex and severe cases, and collectively in its capacity to deal with a more or less large number of cases with various degrees of infection. Last but not least, the epidemic may also importantly depend on the policies and on the calendar of the measures taken at both government and local scales to eradicate the epidemic and on the practical application of these measures.

Once the main variables potentially identified, the question of the governing equations can be another very hard question: how to formulate their dynamical couplings. Numerous model structures have been introduced to consider these quite varied situations. However, these formulations are all based on *a priori* structures. In other words, the processes and their mathematical formulations are assumed to be clearly known which is generally far to be the case, especially when a new disease is concerned. Moreover, once the model formulation has been chosen, a careful parameterization is required. When this parameterization cannot be performed numerically, either due to intrinsic limitations of the observations (inexistent observations of some compartments, poor quality and shortness of the observational dataset, etc.) parameters may have to be borrowed from other contexts or studies (e.g. parameters calibrated on Influenza applied to Covid-19) or chosen based on theoretical considerations (e.g. statistical hypothesis). In such conditions, simulations will be, at best, qualitative. If all the compartments of the model are observed, and data are of sufficiently good quality, a practical calibration may be performed numerically. Unfortunately, it is generally not the case because of the unobserved variables previously mentioned upper in the text. Moreover, under hypothetically ideal situations about the observations, even if all the variables can be observed, an adequate calibration cannot be reached if the governing equations were not chosen adequately. It is therefore necessary to cope with a cross-challenge: to find a model structure and parameterization adapted not only to the proper

dynamical behaviour but also to the observed data set, that is, in few words, adapted to the practical problem.

Various methodologies have been developed to analyse data sets without strong *a priori* knowledge. Artificial Neural Networks, and in particular deep learning, have proven to be very powerful for the modelling in numerous domains. These may provide an alternative approach to the compartment models of pre-fixed structures and to individual based simulations. However, there skills are known to be highly dependent on the training samples and often require an important quantity of data which is rarely available in epidemiology; such data may even be especially scarce as far as a new disease is concerned. Moreover, these approaches generally lead to black box models, that is, efficient predictive models which functioning cannot be assessed. Their ability can thus be very useful to perform forecasts, but relatively poor to understand the dynamical behaviours at work.

2. Theoretical background

2.1. Chaos theory

The theory of nonlinear dynamical systems – or chaos theory – has its origins in the work of the French mathematician Henri Poincaré who was the first one to understand how a system can be both at the same time deterministic (*i.e.* the current state of a studied system at time t entirely determines its immediate future state at time $t + \delta t$) and unpredictable at long term. Chaos theory is very well designed to study real world dynamics because particularly well designed theoretically to study deterministic dynamics that are poorly predictable at long term (Statistical approaches may constitute an alternative to study instable dynamics, but unfortunately abandoning the deterministic part underlying the dynamical behaviour). Striking examples of

such deterministic but unpredictable behaviours could be reproduced numerically in the sixties by Edouard N. Lorenz [1] and Otto E. Rössler [2,3]. René Lozi [4] proved that chaotic behaviours could be produced up to the ultimate precision limits of computers. Contrarily to mathematical approaches – which aim is most often to obtain the analytic solutions from studied equations – chaos theory fosters the use of the phase space (or state space). This orientated space is particularly interesting because it enables a geometric representation of all the states of a dynamical system, independently of time which permits a qualitative representation of the dynamics. Usefully, when the original equations of a studied behaviour are unknown, this space can be reconstructed directly from observational time series. Furthermore, under appropriate conditions, even if several of the system original variables are unknown (or just unavailable), this phase space still can be reconstructed from a subset of observational time series, potentially a single one [5-7]. Finally, to extract a model from observational time series, observational variables should also enable a good observability of the original equations of the underlying dynamic [8,9].

The global modelling technique aims to obtain differential, discrete or delay models from observational time series. It takes its background from statistics and electronic [10,11]. Its differential form was introduced in the early 1990s [12] to obtain models from single time series. Although it has proven early to be applicable to theoretical and experimental time series [13], it is only in the 2000s that it could be applied to observational data from environmental conditions [14,15] and its applications to multiple time series from real environmental conditions is even more recent [16]. The interest to apply this approach can be multiple. It can be used to detect couplings between observational variables [17], to (re-)formulate these couplings algebraically

[18] and, potentially, to understand these couplings based on the reconstructed equations, by inference [16].

2.2 The global modelling technique

In practice, applied to multiple time series, the approach consists in retrieving a set of differential equations in the form $\dot{x}_i = f_i(x_1, x_2, \dots, x_n)$, with $i = 1..n$, where the dots denote the derivatives with respect to time. When variables are missing, for example if only one variable X_1 is available and defined such as $X_1 = h(x_1, \dots, x_n)$ with h the measurement function, then, under proper observability conditions, the original system can be rewritten in the canonical form $\dot{X}_{n-1} = F(X, X_1, \dots, X_{n-1})$ with $X_1 = \dot{X}$, $X_2 = \ddot{X}$, $X_3 = \dddot{X}$, etc. The original system based on variables (x_1, x_2, \dots, x_n) is then reformulated based on the set of variables (X_1, X_2, \dots, X_n) . If the original system of equations (1) is known, under appropriate conditions of observability [8], the canonical form (2) can be derived analytically by using the Lie derivatives

$$L_f h(x_1, \dots, x_n) = \sum_{k=1}^n f_k(x_1, \dots, x_n) \frac{\partial h(x_1, \dots, x_n)}{\partial x_k}, \text{ recursively. The transformation from the original}$$

variables to the variables of the canonical system are thus given by $X_i = L_f^{i-1} h(x_1, \dots, x_n)$.

2.3. Methodology

The GPoM algorithm is used to model and analyse the data set presented in Section 2. The main goal of this algorithm is to obtain equations from observational time series. Therefore, it does not require a strong *a priori* knowledge about the formulation of the equations underlying the dynamics. In practice, to obtain a set of n equations from n observational time series, the following algorithm parameters are required: (1) The maximum polynomial degree q^{\max} , (2) the maximum integration time T^{\max} and (3) the maximum total parameters N_p^{\max} allowed for the

model. For example, considering $n = 3$ observed variables and a quadratic form will lead to 2^m possible formulations for each equation (with $m = 10$), that is 2^{30} potential models. The algorithm is based on the following stages: The first equations is initially considered with all its parameters, (1) a Gram-Schmidt procedure is used to estimate its parameters, (2) a leave-one out method is used to identify the less useful monomial (the one enabling the lower residual reduction), (3) this monomial is removed and the resulting formulation is kept as a potential equation. (4) Steps (1) to (3) are repeated until all the monomial are removed from the equation. Thanks to this procedure, the 2^m possible equations will be reduced to $(m+1)$ potential formulations. The same stages (1) to (4) will be repeated for each equation, leading to $(m+1)^n$ potential model formulations, that is here 11^3 potential models in the present case. The number of model to be tested was thus reduced drastically.

A detailed description of the algorithm can be found in [18] where it is shown that the GPoM algorithm is quite efficient to retrieve the compact formulations of the original systems: in most of the cases, all the terms of the obtained structure are indeed present in the original full formulation although some terms of secondary importance may be missing; fallacious detections are rare. It is also proven that polynomial approximations can be obtained from systems of nonpolynomial form. Note that obtaining a model from a stochastic signal is found to be quite improbable.

3. Organization

The analysis presented in the present study was performed in two steps. In the first step (A) the epidemic was considered at the China scale. The dynamic of the epidemic was modelled using the global modelling technique to investigate the couplings between three observed

variables made available by the Chinese authorities. In a second step (B) the analysis focuses on the number of confirmed cases considering the outbreak in seven provinces in China and two other countries (Japan and South Korea) where the epidemic has been sufficiently long to provide us with dynamical models of canonical form. The obtained models are then used to produce scenarios for fourteen countries where new focuses have broken out. Observational data are used to validate / invalidate the scenarios and to provide diagnostics and prognostics about the current and future evolution of the outbreaks.

Supplementary Material 2-Data correction for country inter-comparison

Two data sets were used to perform the analyses presented in this study: The data from the [National Health Commission of the People's Republic of China \[19\]](#), and the data from the [Johns Hudson University \[20\]](#), both made available from 21 January 2020 to present (20 March 2020).

The first data set was used to study the epidemic at the scale of China. A new methodology was used from 12 February 2020 (included) up to now to count the number of cases. From 12 February 2020 on, persons having developed clinical symptoms of the Covid-19 were included ([\[21\] 13 February 2020](#)), whereas only the cases identified by standard tests based on nucleic acid were counted before. This new methodology aims at providing a complementary picture of the epidemiological situation, enabling an earlier detection of infection cases and thus allowing earlier treatments so as to stop the epidemic. This sudden compensation generated a strong artefact on the daily counts of confirmed cases and deaths on 12 and 13 of February 2020. To take into account these additional cases, this sudden contribution was redistributed among the earlier counts by using a multiplicative factor of 1.3 for the cumulated cases and 1.1 for the cumulated deaths on the corresponding period.

The data from the [Johns Hudson University](#) [20] were used when considering the epidemic at province scale in China, and in other countries. These data were last downloaded on 20 March 2020. Two variables were considered: the daily cumulated number of confirmed cases $C_{\Sigma}(t)$ and deaths $D_{\Sigma}(t)$. A correction of multiplicative factors 1.4 and 1.1 was applied to the Hubei province to account for the counting methodology change already mentioned. Values were missing for numerous European countries (e.g. Italy, Spain, France, United Kingdom, Switzerland, Netherlands, Denmark and Norway) on 13 March 2020 due to a data harmonization at European scale. There were estimated by linear interpolation. The number of cumulated confirmed cases was not updated correctly on 15 March 2020 in France and UK, it was thus replaced similarly.

Obvious differences can be observed between the countries when considering the relationship between the number of cumulated cases and the number of cumulated deaths (Fig. S1). This is obvious when considering the difference between China (in gray) and Italy (in red) (S1a). Several processes can contribute to a differential behaviour, which can result from differences in the distribution of the population in age, gender, social position, on capacity of the health system in the country, etc. However, since a detailed sampling of the population is generally impossible, specific protocols are generally used to sample the population. Large differences may thus result from the protocols used to sample the population. Since the application of this protocol may vary in time and geographically, depending on the situation (saturation in capacity of the health services, limiting stocks of tests, conditions delaying the tests, etc.), it may be difficult to use these original protocols to extrapolate the results at a province or country scale. For this reason, linear regressions were used to inter-calibrate the observations.

The Italian and the Iranian data set were calibrated taking the growing period of Chinese data set as reference, and assuming that the number of deaths was suitable without correction. The correction factor for the number of cumulated cases was estimated to be of 2.45 ± 0.05 for Italy and 1.45 ± 0.05 for Iran (see Fig. S1a). Although the Chinese reference appears quite well adapted when considering very large values of cumulated cases and deaths, its ability to provide a proper reference for small cumulated counts, that is, at the earlier development of the outbreak, appears quite poor. Indeed, the increase of deaths at the earlier development of the outbreak appears twice quicker in comparison to the increase of infection at the earlier period of the record, revealing a clear deficit of infection detection. Assuming that this relationship should have remained linear and constant, the number of early undetected cases can be estimated around 1000 to 2000 cases.

For this reason, the other data sets were calibrated taking the Italian data as direct reference and following the same methodology. Curves before and after calibration are shown in Figures S1b and S1c, respectively. Corresponding coefficients are provided in Table S1.

The calculation of the time derivatives is required to apply the global modelling technique. The daily time series were thus re-sampled at the hourly time step using cubic splines. This process has proven to be a very efficient procedure for the phase space reconstruction and for the global modelling [21]. The computation of the derivatives was then applied using a Savitzky-Golai filter with a ± 4 data points (corresponding to ± 4 hours). The successive derivatives obtained from the original time series $C_{\Sigma}(t)$, $s(t)$ and $D_{\Sigma}(t)$, will be noted $C_k(t)$, $s_k(t)$ and $D_k(t)$, respectively, with k the derivation degree.

Supplementary Material 3-Dynamical regime

Applying a small parameterization change to model **M** (Eq. 1 in main text), a phase non-coherent regime is reached, that is, a behaviour in which several pseudo-frequency are involved. This regime is characterized by much more unpredictable behaviour as it can be observed in Figure S2. After the 15-day transient, the three trajectories diverge much quicker than what was observed in Figure 2 (main text).

Supplementary Material 4-Case Fatality Rate

Model **M** (see Eq. 1 in the main text) was used to estimate the Case Fatality Rate (CFR) of Covid-19. This model was integrated numerically from 21 February 2020 (DoY 52) to 6 September 2020 (DoY 250). The cumulated number of cases $C_{\Sigma}(t)$ and deaths $D_{\Sigma}(t)$ were directly deduced from the daily estimates $C_1(t)$ and $D_1(t)$ (see Fig. S3). The case fatality rate is then directly deduced considering a 5 day delay between the two variables. The analysis shows that the CFR converges to 1.30-1.35 depending on the integration length (from 100 to 230 days). Changing the dynamical regime was found to be marginal. This value higher in comparison to previous estimates ([22] 5 March 2020), but more than twice smaller than the estimates directly deduced from the Chinese data sets from Wuhan suburb to China scale 3.97-4.94 (see Table S2).

Supplementary Material 5-Canonical models

The global modelling technique was used in the present analysis considering two model general structures. The first structure was used to retrieve the couplings between variables $C_\Sigma(t)$, $s(t)$ and $D_\Sigma(t)$ (see [Eq. \(1\)](#) in the main text).

5 For the other analyses, a structure of canonical form (see [Suppl. Mat. 1](#)) was used to retrieve a multidimensional model from single time series. The algebraic structures and phase portraits of the models used to perform scenarios are presented hereafter. The following model functions were obtained, for the Hubei province

$$10 \quad F_{\text{Hubei}}(C_\Sigma, C_1, C_2) = 728.374334674 + 0.12025655855 C_2 - 2.44374861116086 C_1 \\ + 6.115900961 C_1 C_2 + 0.67320557758 C_\Sigma - 6.7065922695 \cdot 10^{-6} C_\Sigma C_2 \\ - 2.14395055039 \cdot 10^{-5} C_\Sigma C_1 - 8.42187611295 \cdot 10^{-6} C_\Sigma^2, \quad (\text{Eq. S1})$$

which projection (C_Σ, C_1) of the phase portraits is given in [Figure S4a](#), with initial conditions (617.0938896, 146.4350011, 116.6699932); For the Zhejiang:

$$15 \quad F_{\text{Zhejiang}}(C_\Sigma, C_1, C_2) = 33.02877246 + 0.72175282909 C_2 - 3.0997963808 C_1 \\ + 0.01071494081 C_1^2 + 0.686737990591491 C_\Sigma - 0.00136034943 C_\Sigma C_2 \\ - 0.00070734290234 C_\Sigma C_1 - 0.00068054968885 C_\Sigma^2, \quad (\text{Eq. S2})$$

which projection (C_Σ, C_1) of the phase portraits is given in [Figure S4b](#) with initial conditions (26.039364, 14.6265229, -1.089333); For the Henan:

$$\begin{aligned}
F_{\text{Henan}}(C_{\Sigma}, C_1, C_2) = & 15.7875118 + 0.037323657C_2^2 - 2.16550311440939C_1 \\
& - 0.011262814729C_1C_2 - 0.00655453369C_1^2 + 0.6106396255C_{\Sigma} \\
& - 0.00068928413C_{\Sigma}C_1 - 0.000955568338357319C_{\Sigma}^2,
\end{aligned} \tag{Eq. S3}$$

which projection (C_{Σ}, C_1) of the phase portraits is given in [Figure S4c](#) with initial conditions (2.6233721, 0.3184398, 3.27712507); For the Jiangxi:

5

$$\begin{aligned}
F_{\text{Jiangxi}}(C_{\Sigma}, C_1, C_2) = & 32.204549591 + 0.49031399622C_2 - 3.1140961152C_1 \\
& + 0.002199220326C_1^2 + 0.68989562079C_{\Sigma} - 0.0013919594626C_{\Sigma}C_2 \\
& + 0.00026427097753C_{\Sigma}C_1 - 0.00071476276503C_{\Sigma}^2,
\end{aligned} \tag{Eq. S4}$$

which projection (C_{Σ}, C_1) of the phase portraits is given in [Figure S4d](#) with initial conditions (9.93751978, 14.1311195, 1.9752268239); For the Guangdong:

10

$$\begin{aligned}
F_{\text{Guangdong}}(C_{\Sigma}, C_1, C_2) = & -1.02106445487 + 0.81072130655C_2 - 1.395924231C_1 \\
& - 0.012327743251C_1C_2 + 0.543560087856C_{\Sigma} - 0.0016197665499C_{\Sigma}C_2 \\
& - 0.0020056527991C_{\Sigma}C_1 - 0.00090789371754C_{\Sigma}^2,
\end{aligned} \tag{Eq. S5}$$

which projection (C_{Σ}, C_1) of the phase portraits is given in [Figure S4e](#) with initial conditions (15.36734533, 8.0091208674, 5.6142410724); For the Anhui:

15

$$\begin{aligned}
F_{\text{Anhui}}(C_{\Sigma}, C_1, C_2) = & 31.38290829 + 0.38390685661C_2 - 3.5311710824C_1 \\
& - 0.00731724814919C_1^2 + 0.774413823641C_{\Sigma} - 0.0012193215228C_{\Sigma}C_2 \\
& + 0.001313318079C_{\Sigma}C_1 - 0.0010396400224C_{\Sigma}^2,
\end{aligned} \tag{Eq. S6}$$

which projection (C_Σ, C_1) of the phase portraits is given in [Figure S4f](#) with initial conditions (7.3752626332, 1.9460686395, 1.8400134404); For the Heilongjiang:

$$\begin{aligned}
 F_{\text{Heilongjiang}}(C_\Sigma, C_1, C_2) = & -0.45216004565 C_2 - 1.7938406094 C_1 + 0.02704335177 C_1 C_2 \\
 & - 0.0050750328249 C_1^2 + 0.68013328873 C_\Sigma - 0.0012900540558 C_\Sigma C_2 \\
 & - 0.00178976611992496 C_\Sigma C_1 - 0.00106038366348117 C_\Sigma^2,
 \end{aligned}
 \tag{Eq. S7}$$

5

which projection (C_Σ, C_1) of the phase portraits is given in [Figure S4g](#) with initial conditions (2.9464454755, 1.8801103178, 0.80343957214); And for South Korea:

$$\begin{aligned}
 F_{\text{South Korea}}(C_\Sigma, C_1, C_2) = & -2.3996225057 C_1 + 0.0025273468239 C_1 C_2 \\
 & - 0.0012812217959 C_1^2 + 1.0858468259 C_\Sigma - 0.00038068284318 C_\Sigma C_2 \\
 & + 3.1133342825 \cdot 10^{-5} C_\Sigma C_1 - 0.00014594323375 C_\Sigma^2,
 \end{aligned}
 \tag{Eq. S8}$$

10

which projection (C_Σ, C_1) of the phase portraits is given in [Figure S4h](#) with initial conditions (1.1440650785, 1.0713288594, 0.53117662272).

The comparison between models and observations are shown in [Figures S4i-j](#) for cumulative cases S_Σ and [S4k-l](#) for daily new cases S_1 . Basic statistics for model validation are presented in [Tables S3](#) and [S4](#) for S_Σ and S_1 , respectively.

15

Supplementary Material 6-Italy Covid-19 model

Taking into account the last data made available today 26 March 2020 from the Johns Hopkins University (*i.e.* time series up to 25 March 2020), a canonical model could also be obtained for Italy. Note that, as for the other models, *no* strong a priori structure was used to

20

constrain the model: it was entirely derived from the observational time series, without strong hypotheses. In other words, it is not just a calibration: the dynamic is not assumed to be known. This is an important element to have in mind since such a general polynomial structure can generate a huge diversity of dynamics, such as: highly dissipative chaotic dynamics (e.g. the Lorenz-1963 [1] and Rössler-76 [2] systems), weakly dissipative dynamics (cereal crops attractor [15]), but also SEIR epidemiological models, quasiperiodic and periodic behaviours, among which, numerous of them can be derived analytically [23]. Moreover, contrarily to most of the modelling approaches commonly used, the chaos based approach here used is also independent to the initial conditions: the equations obtained here are deduced from the observed data set but are not specific to the observed initial conditions.

For the first time since the beginning of the outbreak in Italy, a numerically integrable model was obtained. Simulations are shown in [Figure S7](#). These simulations suggest that the maximum of the epidemic has now been reached and the ending stages can now be estimated for the Italian outbreak. Although new clusters cannot be excluded yet, obtaining such a model clearly shows that the main current focuses of Covid-19 in Italy are now stabilizing. Among the four models, models M_{I11} and M_{I12} can only marginally reproduce the oscillations observed in the original data set and are thus very likely to oversimplify the dynamic. Models M_{I13} and M_{I14} appears more reliable. Based on their simulations, it is estimated that the cumulative number of cases may reach 243-263 thousand cases (98-106 thousands before correction). Based on this model, the ending stages can also be estimated. Still based on models M_{I13} and M_{I14} , the stage corresponding to a level of ~10 new cases per day is estimated between day of year 130 and 144.

Supplementary Material 7-Comparative analysis

Various modelling approaches have been developed to formulate scenarios with the aim to provide well-grounded arguments for decision making. In particular, for the present Covid-19 pandemic, an Agent Based model initially developed to support pandemic influenza planning, was used to provide a comparison of suppression versus mitigation strategies under real pandemic situation [24]. Such models can simulate the behaviours at various scales, both in time and space and was applied to the UK and the US. Such type of model can include numerous variables and enable realistic simulations. In practice, such models require the identification of the main active agents, and the formulation – at the agents scale – of the main processes at work in the propagation of the disease. These obvious interests will also have some specific limitations. In particular, the main agents and agent properties described in the model will have to be known, sufficient to provide a complete description of the situation (up to a certain degree of approximation), valid in their formulation, and properly calibrated. Each of these points may meet important difficulties. Essential agents may not always be known (e.g. for the present pandemic: the original source of the contamination, the case-0 of each outbreak, the main vectors, etc.), rough approximations may have to be used to represent the complexity of the situation (for numeric reasons, linear assumptions may have to be fostered, retroactions may be avoided or defined with an extreme caution), data may be missing to calibrate such complicated models (many variables at work, even if under linear approximation, mostly) and numerous assumptions may be required (e.g. locations where the contacts between the agents can take place, parameterization of the infection distance, incubation period, level of infectiousness depending on the agent status, inter-country propagation growth rate, etc.). Finally, the large number of agents and variables involved in this type of model will also require a high

computation cost, which may lead to delay the reaction, and to make the simulation potentially less understandable, and their interpretation possibly more subjective.

The approach used in the present study takes the opposite point of view. Instead of assuming the agents (or the variables) and the processes at work to be known, its aim is to extract the governing equations from the observations, without any strong a priori hypotheses. Its advantages and drawbacks may thus completely differ and it may thus potentially offer an independent validation while leading to useful discussion otherwise.

The global modelling technique has never been used under operational conditions: the model for the Bombay plague was obtained one century after the Bombay epidemic took place, and the model for the West Africa outbreak of Ebola Virus Disease, although obtained during the outbreak was published after the outbreak was over. It is thus the first time that this technique is investigated to test its potential for decision making. To do so, the global modelling approach is not used simply as a modelling approach but to perform inter comparisons between primary clusters in Chinese provinces and various countries in Asia and Europe. As it is presently, it cannot be used to establish specific scenarios but to identify the most relevant scenarios among the situations already experienced in other focuses. In other words, it does not have the same potential as the other approaches, but it may potentially bring information that other approaches cannot have.

References:

1. **Lorenz EN**, Deterministic non-periodic flow. *Journal of Atmospheric Science* 1963, 20(2): 130–141

2. **Rössler OE**, An Equation for Continuous Chaos, *Physics Letters* 1976, **57A**(5), 397-398.
3. **Rössler OE**, An equation for hyperchaos, *Physics Letters A* 1979, **71**, 155-157.
4. **Lozi R**, Giga-periodic orbits for weakly coupled tent and logistic discretized maps. In:
Siddiqi AH, Duff IS, Christensen O eds. *Proceedings Conference International on*
5 *Industrial and Applied Mathematics* 2007, Modern mathematical models, methods and
algorithms for real world systems, New Delhi, 4–6 Dec. 2004, Anamaya Publisher (New
Delhi), 80-14.
5. **Whitney H**, Differentiable manifolds, *Annals of Mathematics*, **37**(3), 645-680, 1936.
6. **Takens F**, *Detecting strange attractors in turbulence. In: Rand D.A. and Young L.-S.*
10 *(eds.). Dynamical Systems and Turbulence, Lecture Notes in Mathematics* 1981, **898**,
Springer-Verlag, 366-381.
7. **Sauer T, Yorke JA, Casdagli M**, Embedology, *Journal of statistical Physics* 1991,
65(3-4), 579-616.
8. **Aguirre LA, Letellier C**, Investigating observability properties from data in nonlinear
15 dynamics. *Physical Review E* 2011; **83**, 066209.
9. **Letellier C et al.**, A symbolic network-based nonlinear theory for dynamical systems
observability, *Scientific Report* 2018, **8**, 3785, DOI:10.1038/s41598-018-21967-w
10. **Billings SA**, Structure detection and model validity tests in the identification of nonlinear
systems, *ACSE Report* 1982, **196**, 22.
- 20 11. **Aguirre LA, Billings SA**, Retrieving dynamical invariants from chaotic data using
NARMAX models, *International Journal of Bifurcation and Chaos* 1995, **9**(2), 449-474.

12. **Gouesbet G, Letellier C**, Global vector-field reconstruction by using a multivariate polynomial L2 approximation on nets. *Physical Review E* 1994, **49**(6), 4955-4972.
13. **Letellier C et al.**, Topological Characterization and Global Vector Field Reconstruction from an experimental electrochemical system, *Journal of Physical Chemistry* 1995, **99**,
5 7016-7027.
14. **Maquet J, Letellier C, Aguirre LA**, Global models from Canadian lynx cycles as a direct evidence for chaos in real ecosystems. *Journal of Mathematical Biology* 2007, **55**(1), 21-39.
15. **Mangiarotti S et al.**, Polynomial search and global modeling: two algorithms for
10 modeling chaos. *Physical Review E* 2012, **86**(4), 046205.
16. **Mangiarotti S**, Low dimensional chaotic models for the plague epidemic in Bombay (1896–1911), *Chaos, Solitons & Fractals* 2015, **81**, 184-196.
17. **Mangiarotti S, Sendiña-Nadal I, Letellier C**, Using global modelling to unveil hidden couplings in small network motifs, *Chaos* 2018, **28**, 123110.
- 15 18. **Mangiarotti & Huc**, Can the original equations of a dynamical system be retrieved from observational time series? *Chaos* 2019, **29**, 023133. doi:10.1063/1.5081448.
19. **National Health Commission of the People's Republic of China** (2020).
http://www.nhc.gov.cn/yjb/pzhgli/new_list.shtml (Last assessed on 21 March 2020)
20. **Johns Hudson University** (JHU 2020). https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data (last assessed on 21 March 2020)
- 20 21. **Mangiarotti S, Zhang Y, Leblanc M**, Is it possible to erroneously detect low-dimensional determinism and chaos from stochastic signals using the global modelling

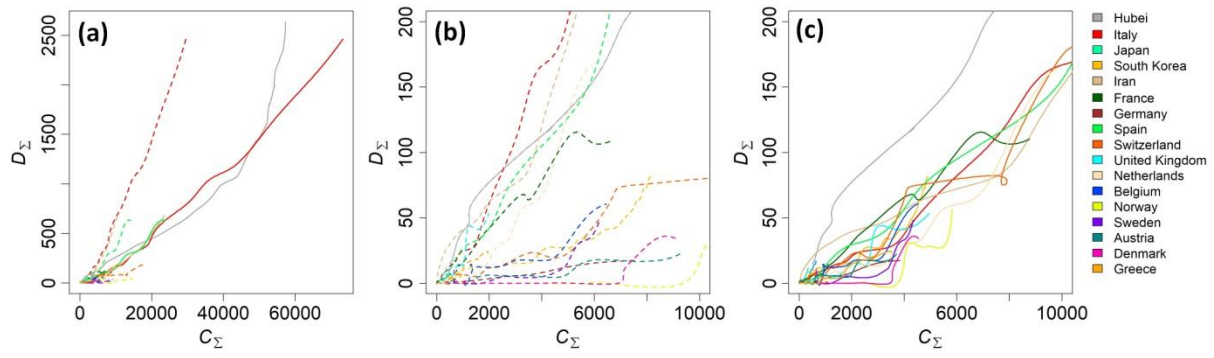
technique? Electronic Supplementary Material ESM1, Chaos theory applied to the modelling of karst springs: first results from univariate time series, *Hydrogology Journal* 2019, **27**, 2027-2043. <https://doi.org/10.1007/s10040-019-01971-8>

22. **World Health Organization**. Coronavirus disease (COVID-2019) situation report—30.
5 https://www.who.int/docs/default-source/coronaviruse/situationreports/20200219-sitrep-30-covid-19.pdf?sfvrsn=3346b04f_2 (accessed 5 March 2020).

23. **Lainscsek C, Letellier C, Schürerer C**, Ansatz library for global modeling with a structure selection, *Physical Review E* 2001, **64**, 016206.

24. **Ferguson et al.** (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce
10 COVID19 mortality and healthcare demand. DOI: <https://doi.org/10.25561/77482>

Fig. S1. Relation between Cumulative cases and cumulated deaths



5

Figure S1: Relation between Cumulative cases C_{Σ} and cumulated deaths D_{Σ} for Hubei and 16 countries on their full range of values (a) and on selected ranges (b-c). Dashed lines and plain lines refer to uncorrected and corrected cumulated cases, respectively.

Fig. S2. Observed and modelled time series under phase non-coherent regime

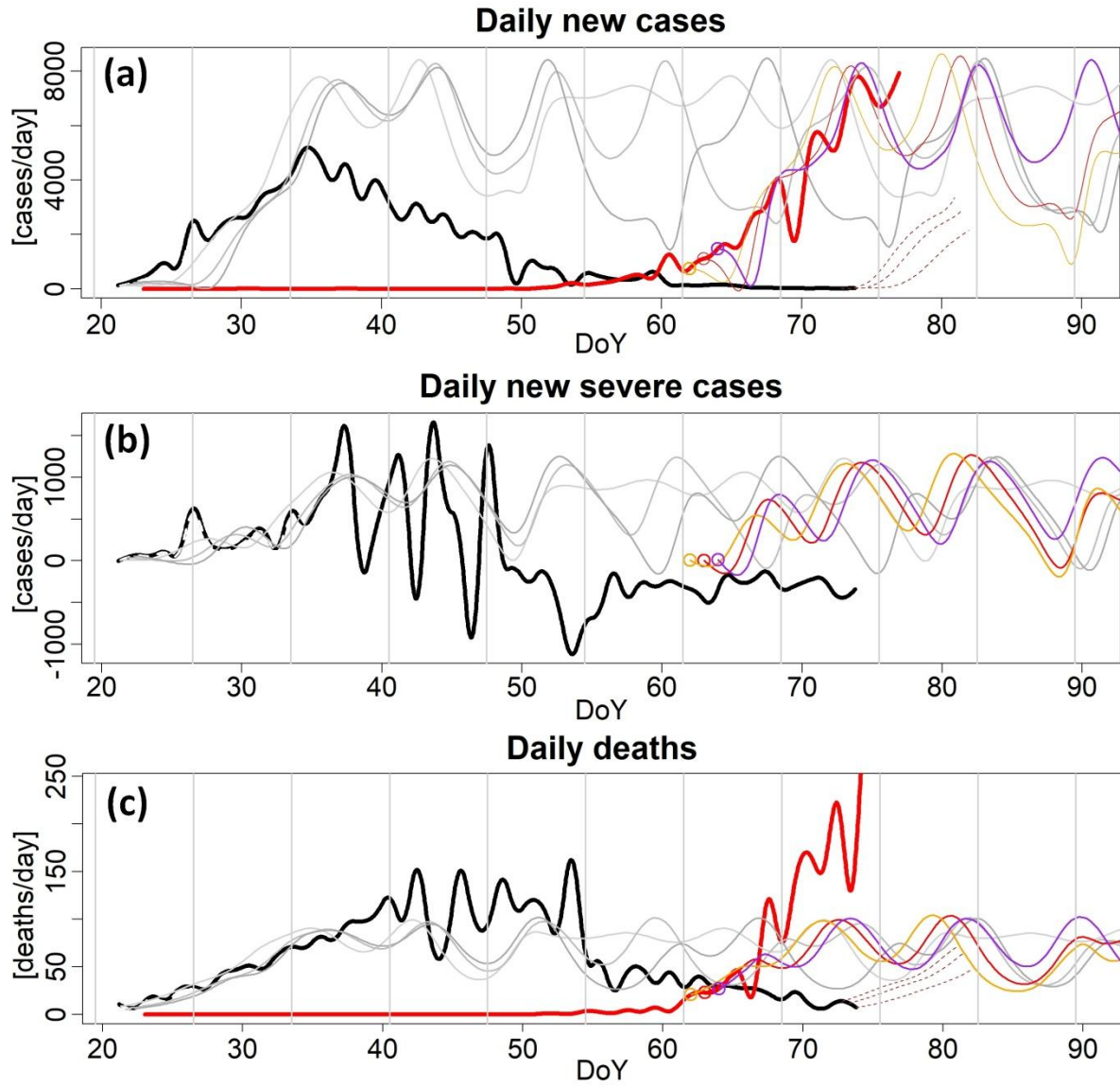
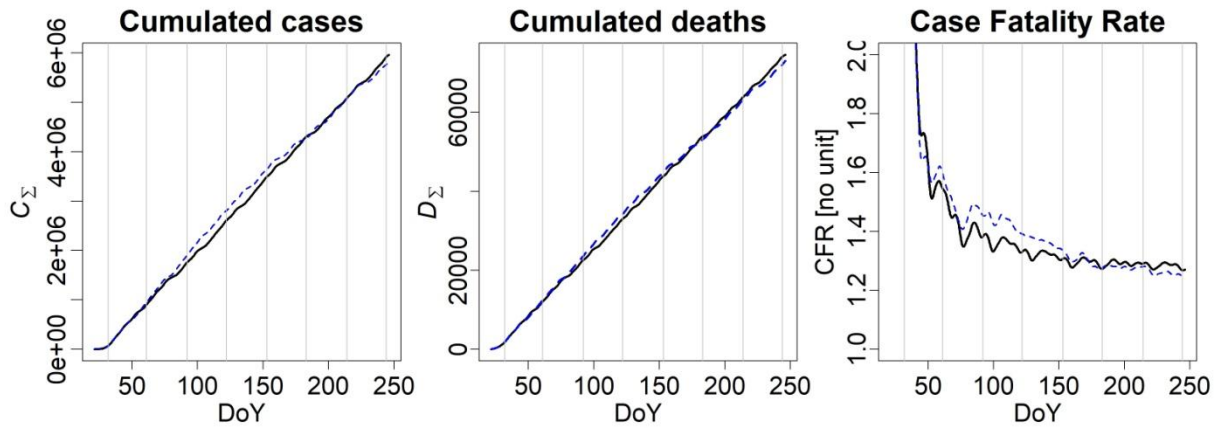


Figure S2: Same as **Figure 2** (in the main text) but with a slight parameter change of one monomial, that is $+0.154574 s_1 D_1$ instead of $+0.15204 s_1 D_1$. The regime becomes even more chaotic: the topological structure of the updated chaotic attractor involves several foldings instead of a single one before, following the same characteristics as the Rössler-1976 chaotic system[2]. A more detailed topological analysis remains to be performed.

5

Fig. S3. Case Fatality Rate



5

Figure S3: Cumulated cases (left panel), cumulated deaths (middle panel) and Case Fatality Rate (right panel) deduced from model **M** (Eq. 1 in the main text). Two simulations are provided, one based on a phase-coherent chaotic regime (black), another one based on a phase-non coherent chaotic regime (blue).

Fig. S4. Canonical models validation

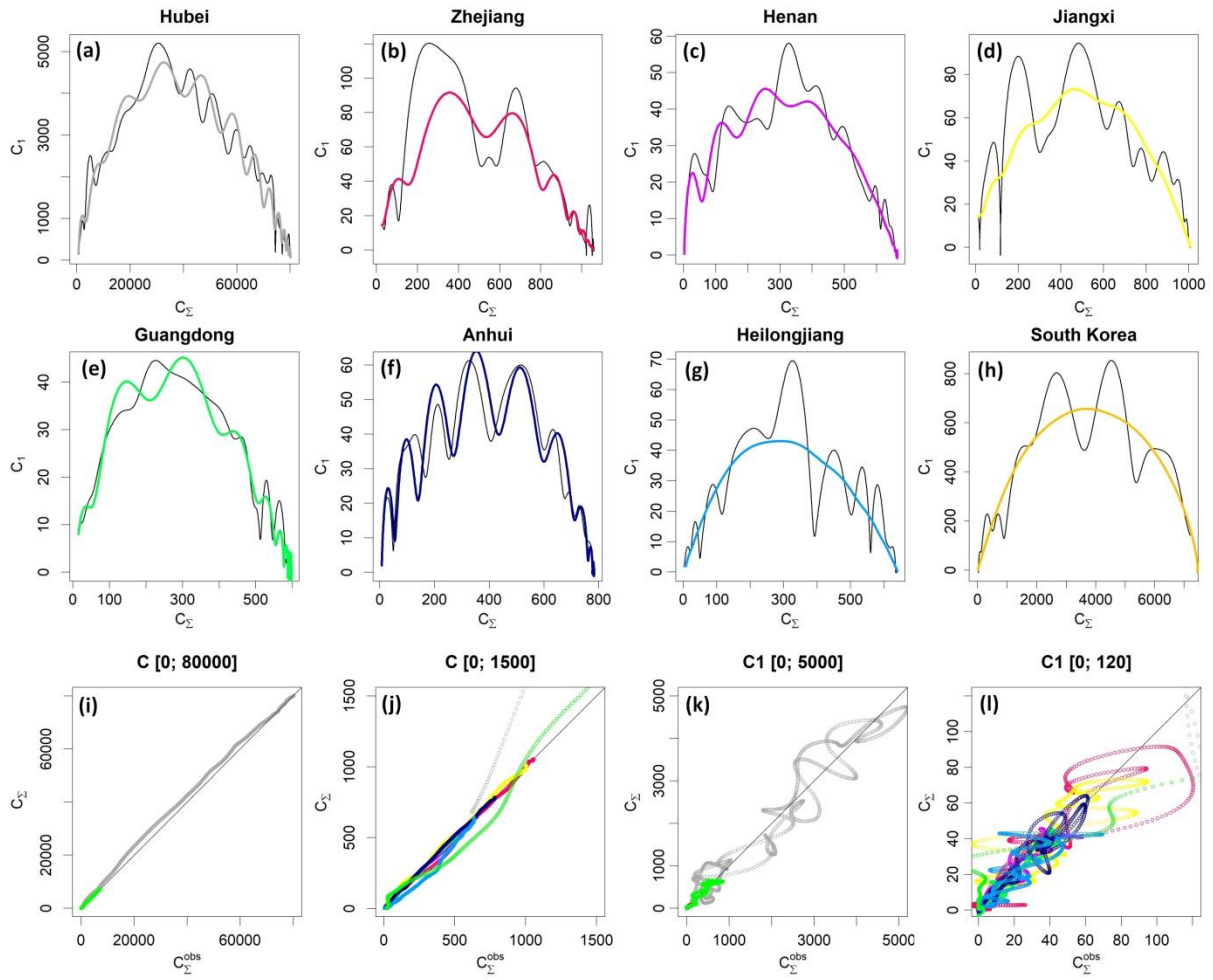


Figure S4: Canonical models validation. Original (in colour) and modeled (in black) phase portraits in (C_Σ, C_1) projection for Hubei (a), Zhejiang (b), Henan (c), Jiangxi (d), Gouangdong (e), Anhui (f) and South Korea (h). Scatter plots – modelled \hat{C}_Σ versus observed C_Σ^{obs} in (i-j) and modelled \hat{C}_1 versus observed C_1^{obs} in (k-l). For more details, see **Suppl. Mat. 5**.

Fig. S5. Empirical scenarios simulations (variable C_1)

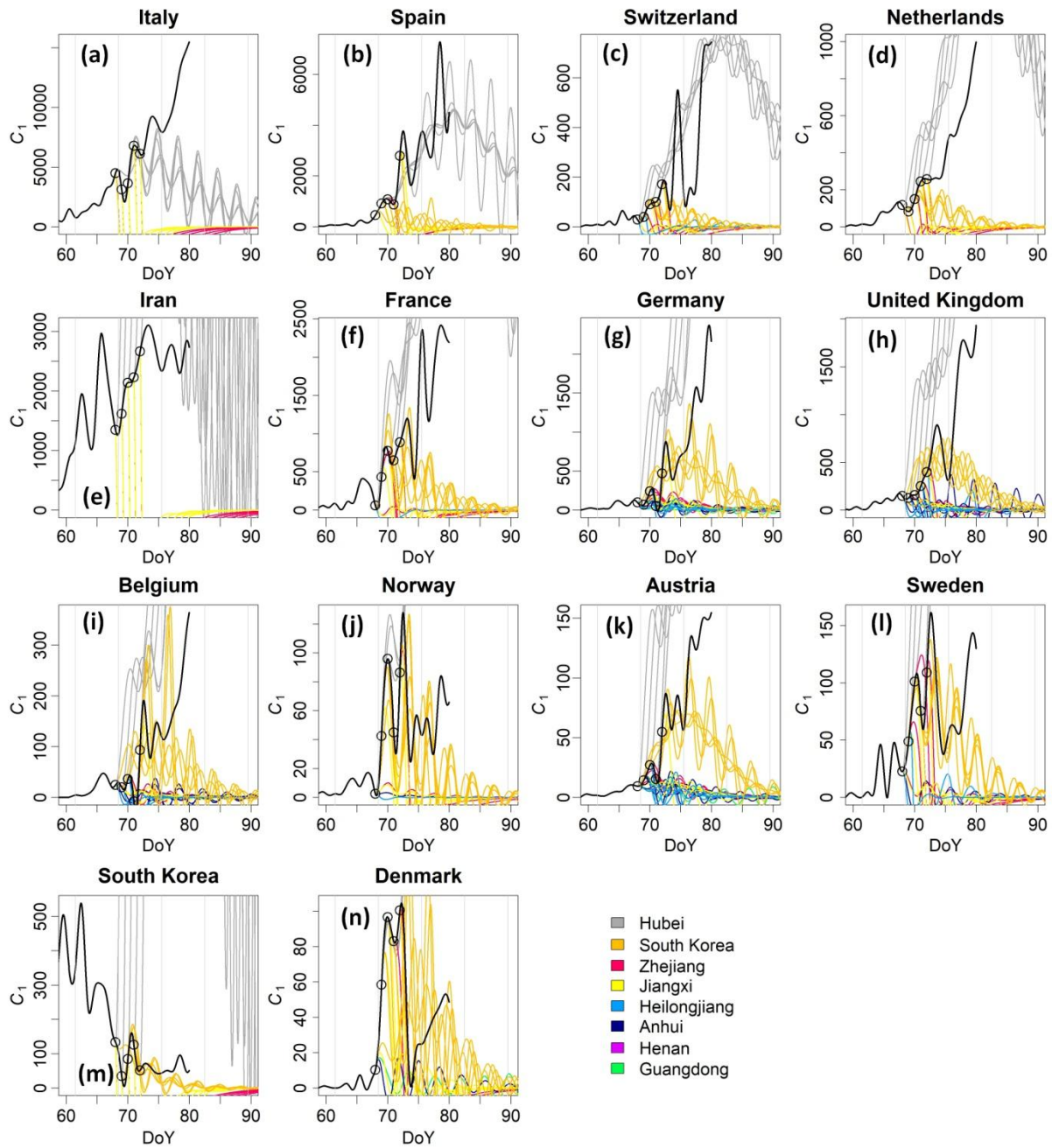


Figure S5: Epidemic curves $C_1(t)$ resulting from the scenarios for fourteen countries based on the models obtained from seven Chinese provinces and South Korea. For each model, an ensemble of five simulations was run starting from the observational initial conditions (black circle) 7 March (DoY 69) to 11 March (DoY 72) 2020. Correction factors used to make the country inter comparison are provided in **Suppl. Mat. 2**. The Chinese data set is used as reference. Observations are in black.

5

Fig. S6. Epidemic curves and control

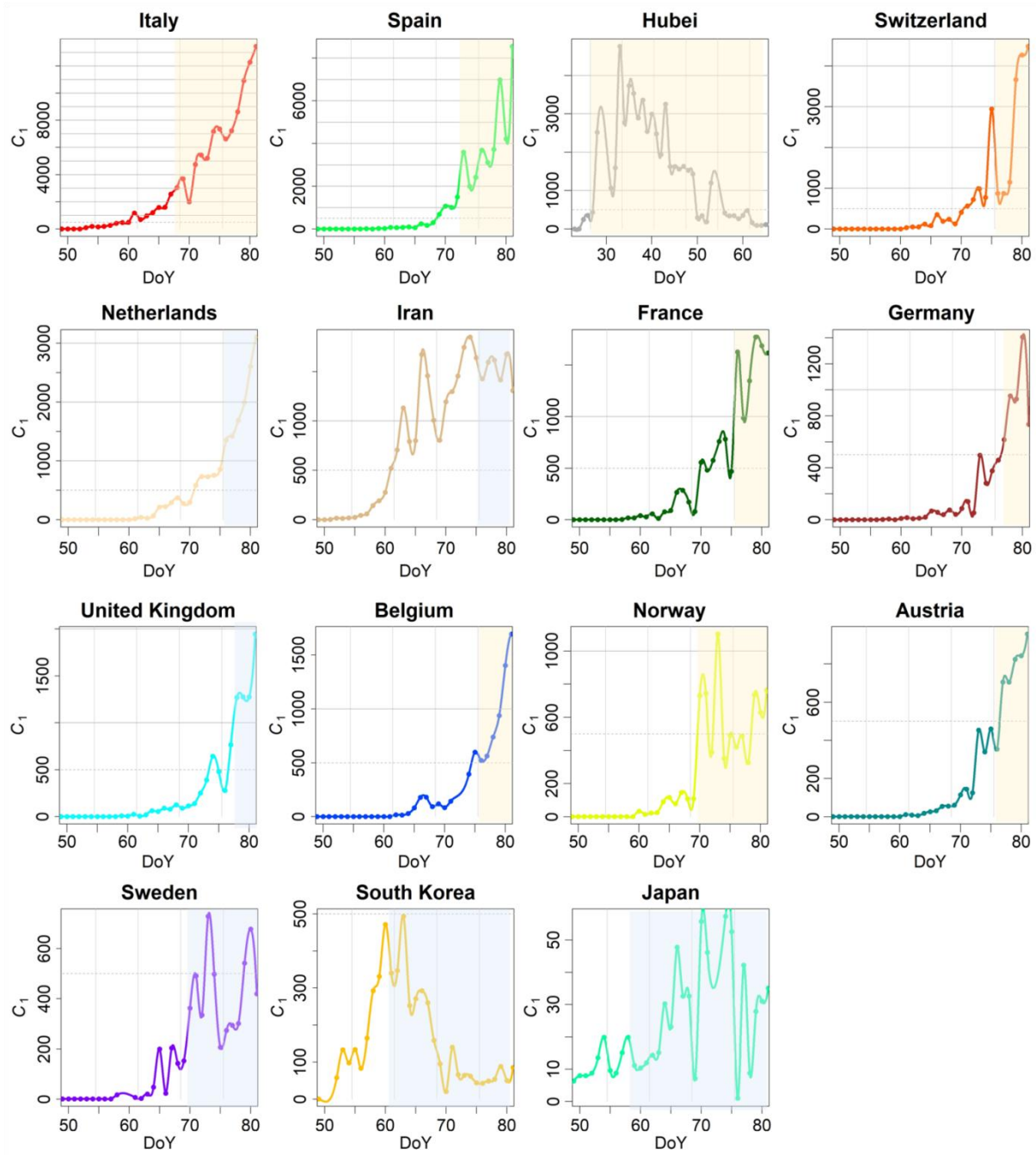


Figure S6: Epidemic curves $C_1(t)$ for Hubei province in China and fourteen countries in Asia and Europe. Period with most stringent measures taken by the authorities are shadowed; the level of enforcement of nationwide lockdown varies between countries: from strict state control (yellow) to advice given to citizen and volunteer basis (blue).

Fig. S7. Observed and modelled Covid-19 outbreak in Italy

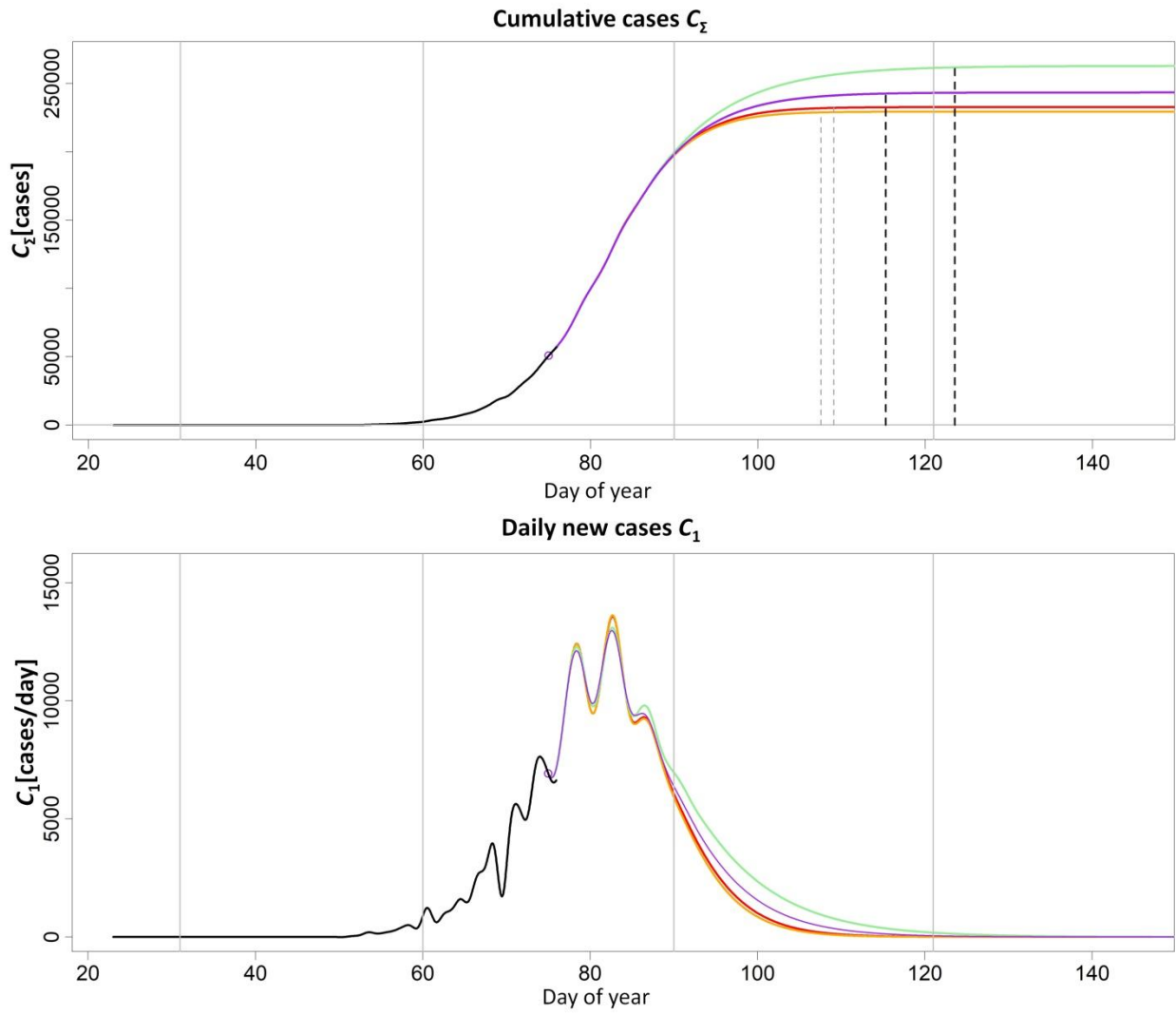


Figure S7: Observed (black line) and Modelled (color lines) cumulative counts of infectious cases $C_{\Sigma}(t)$ (top panel), and daily new cases $C_1(t)$ (bottom panel). The four simulations correspond to models M_{I1} (orange), M_{I2} (red), M_{I3} (green), M_{I4} (purple) of increasing complexity (7 to 10 terms in the polynomial function). Mode details are provided in **Suppl. Mat. 6**. Dates corresponding to $C_1^{\max}/100$ (approximately 130 new cases per day) are denoted by dashed lines, and respectively correspond to day of year 109, 107, 124 and 115. A more detailed information is provided in **Table S5**.

5

Table S1. Correction coefficients

Country	China	United Kingdom	Italy	Iran	Spain	Netherlands	France	Japan
Coeff.	1.	2.5±0.2	2.48±0.05	2.25±0.05	1.7±0.1	1.7±0.2	1.3±0.2	2.1±0.5
	Belgium	Sweden	Switzerland	South Korea	Germany	Denmark	Austria	Norway
	0.7±0.2	0.7±0.2	0.6±0.2	0.6±0.1	0.5±0.2	0.5±0.2	0.4±0.2	0.4±0.2

5 **Table S1:** Correction coefficients (n.u.) applied to the cumulated cases, country by country. The China data set is taken as reference. Coefficients vary from 2.5 in United Kingdom to 0.4 in Austria and Norway. Coefficients greater than one suggest a deficit of sampling in comparison to China whereas coefficients lower than one reveal an improved sampling. Assuming that the sampling performed by Norway is not biased would mean that all the estimates, in the analysis,
10 should be multiplied by 2.5.

Table S2. Case fatality rate and mortality

	Population (in million)	Total confirmed cases	Total deaths	Case Fatality Rate (in %)	Mortality Rate (in ‰)
China	1427.65	81032	3217	3.97	0.00225
Hubei	59.17	67798	3099	4.57	0.0524
Wuhan suburb	11.08	50003	2469	4.94	0.2228

5

Table S2: Case fatality rate and mortality rate resulting from Covid-19 epidemic at China, Hubei and Wuhan suburb scales and estimated on 21 March 2020 (one relatively small part of infected people ~1000 has not recovered yet).

10

Table S3. Model validation for cumulative cases

C_{Σ}	Hubei	Zhejiang	Henan	Jiangxi	Gouangdong	Anhui	Heilongjiang	South Korea
A	0.985	0.997	0.994	0.977	0.989	0.989	1.029	1.007
B	2151.	-1.778	0.6271	22.494	7.6090	10.06	-40.25	-8.272
R^2	0.999	0.997	0.999	0.999	0.999	0.999	0.989	0.999
p -value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

5 **Table S3:** Statistics for the validation of the canonical models based on the cumulated number of infection C_{Σ} .

Table S4. Model validation for daily new cases

C_1	Hubei	Zhejiang	Henan	Jiangxi	Gouangdong	Anhui	Heilongjiang	South Korea
A	0.9466	0.801	0.908	0.830	0.978	0.976	0.755	0.906
B	101.3	4.788	1.514	4.104	0.232	0.449	3.807	24.73
R^2	0.923	0.837	0.919	0.844	0.950	0.943	0.721	0.901
p -value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

5 **Table S4:** Statistics for the validation of the canonical models based on the daily number of new infection C_1 .

Table S5. Ending stages

Models	Stage date (in Day of year)				C_{Σ}^{\max}	
	$C_1 = 130$	$C_1 = 100$	$C_1 = 10$	$C_1 < 1$	Corrected	Uncorrected
M_{I1}	109	111	121	131	233000	94000
M_{I2}	107	109	118	127	230000	92000
M_{I3}	124	126	144	162	263000	106000
M_{I4}	115	117	131	145	243000	98000

- 5 **Table S5:** Ending stages (in Day of year) estimated with the four Italian models for various values of daily new cases C_1 . Estimated maximum cumulative counts of infectious cases C_{Σ}^{\max} ; both corrected and uncorrected values are provided to make the comparison with the official results easier.