International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:15, No:1, 2021

The Origin, Diffusion and a Comparison of Ordinary Differential Equations Numerical Solutions Used by SIR Model in Order to Predict SARS-CoV-2 in Nordic Countries

Gleda Kutrolli, Maksi Kutrolli, Etjon Meco

Abstract—SARS-CoV-2 virus is currently one of the most infectious pathogens for humans. It started in China at the end of 2019 and now it is spread in all over the world. The origin and diffusion of the SARS-CoV-2 epidemic, is analysed based on the discussion of viral phylogeny theory. With the aim of understanding the spread of infection in the affected countries, it is crucial to modelize the spread of the virus and simulate its activity. In this paper, the prediction of coronavirus outbreak is done by using SIR model without vital dynamics, applying different numerical technique solving ordinary differential equations (ODEs). We find out that ABM and MRT methods perform better than other techniques and that the activity of the virus will decrease in April but it never cease (for some time the activity will remain low) and the next cycle will start in the middle July 2020 for Norway and Denmark, and October 2020 for Sweden, and September for Finland.

Keywords—Forecasting, ordinary differential equations, SARS-CoV-2 epidemic, SIR model.

I. Introduction

THE outbreak of the new coronavirus infection is just before and after the Chinese Spring Festival (officially until now). So this epidemic started in early December 2019. On 12 January, the World Health Organization (WHO) confirmed that a novel coronavirus was the cause of a respiratory illness in a group of people in Wuhan, in Hubei province, China. The Health Commission of Wuhan, Hubei, informed WHO about the situation on 31 December 2019. After that, Nordic countries issued a press release highlighting the discovery of the novel coronavirus, and their monitoring of the situation. The risk of spread to Nordic countries was described as ”very low” as there was yet no evidence that the virus could spread between humans but it was recommended to avoid unnecessary travel to Hubei province.

After the WHO classified the novel coronavirus as a Public Health Emergency of International Concern on 30 January and demanded that all member states should cooperate to prevent further spread of the virus, Nordic countries started to classify the novel disease as a notifiable infectious disease which would be dangerous to public health and dangerous to society.

G. Kutrolli is with the Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy (e-mail: g.kutrolli1@campus.unimib.it).
M. Kutrolli is with Smart-Tech and Department of Computer Science, University of Aleksander Moisiu, Albania.
E. Meco is with Cell Department, Toxicon Corporation, USA.

The very first cases in Nordic countries started at the end of January (on 29 January for Finland, 31 January for Sweden, 26 February for Norway, and 27 February for Denmark) and the number started to increase from one day to another until the countries announced lockdowns one by one. The lockdown in Norway was announced on 12 March 2020 and on 13 March 2020, Denmark announced a lockdown as well followed by Finland on 16 March 2020. Sweden, unlike many other countries, has not imposed any lockdown. The measures Norway, Finland and Denmark took were in line with those introduced in other European countries such as Italy. In all Nordic countries under study, there are 34418 confirmed infected cases, 2458 deaths and 7471 recovered until 21 April 2020 (reported in WHO). A more general view about the current situation of the coronavirus from the beginning of the spread until 21 April 2020 is given in Fig. 1, Fig. 2 and Fig. 3. From these figures it is clear to see that Sweden has arrived in an epidemic size larger than other Nordic countries, while the country with lowest number of infected people is Finland. Referring to Fig. 2, Sweden is at the top list of the Nordic countries having the highest number of deaths. Finland and Norway have the lowest number of deaths, followed by Denmark. Regarding the number of recoveries the highest number belongs to Denmark. All of these numbers reflect the preventative measures that governments took since in the beginning of the epidemic spread, which explains why Sweden has the highest number of infected cases and deaths. In our analysis, we are concerned with the study of the spread of the coronavirus from the origin, its diffusion and predicting the situation in the nearest future. For this reason, we include cases of China and Italy.

Since the effective and well-tested vaccine and medicine against SARS-CoV-2 have not been invented, a key part in managing this pandemic is flattening t he e pidemic curve through prevented measures undertaken by governments. At this point, data scientists and researchers play a crucial role helping governments to take the right decisions by their contribution in studying and understanding better the virus and its diffusion. Many researchers proposed and analyzed different models to estimate the spread of other diseases such as Ebola and SARS. In [9], the author studies three types of time series (AR(1), ARIMA(0,1,0) and ARMA(1,1)) in modeling and forecasting the SARS epidemic in China. The differences between the observed and forecasts from AR could

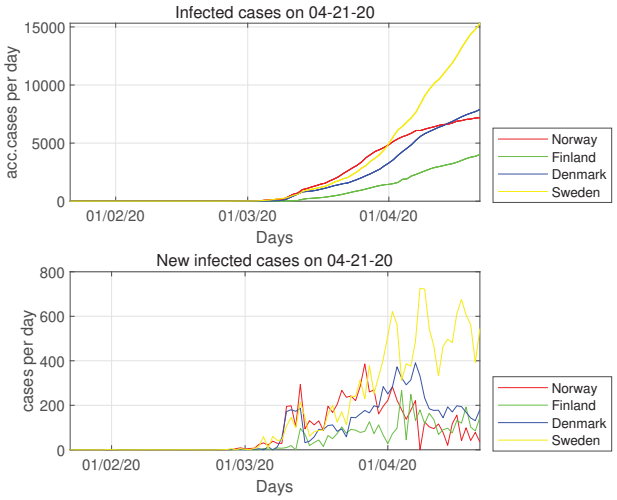

Fig. 1 The progression of the virus with infected cases over time

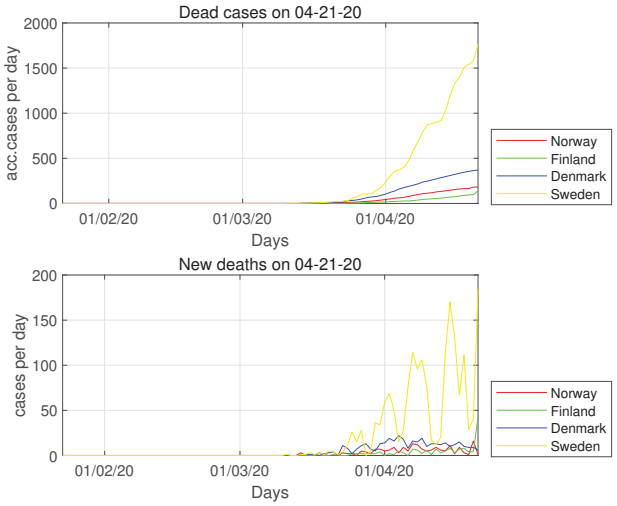

Fig. 2 The progression of the virus with deaths over time

be used as a measure of the effectiveness of the interventions. ARIMA is mainly used for short term forecasting, as well as ARMA is used to assess the relative long-term effect of SARS from a cumulative point of view. In [7], the authors construct dynamic Ebola virus disease transmission models to predict epidemic trends and evaluate intervention measure efficacy f ollowing t he 2 014 E bola e pidemic i n W est Africa. Using a (Susceptible, Infectious, Recovered) SIR model, it is found that the population would be dramatically decreased in a short time in the absence of control and prevention measures. The SIR model indicated that the Susceptible, Infectious and Recovered population fluctuate p eriodically a round their dynamic equilibrium points and that these results support the idea that the Ebola epidemic may break out periodically, though the disease will reach a dynamic equilibrium point with a shorter period between epidemics of progressively weaker intensity. In [2], the author uses a (Susceptible,

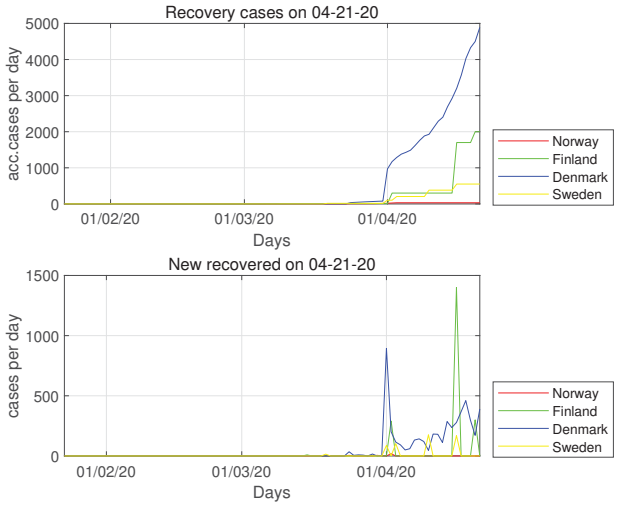

Fig. 3 The progression of the virus with recovered cases over time

Exposed, Infectious, Recovered) SEIR model to simulate the transmission of the Ebola disease. Moreover, it is implemented an optimal control in SEIR model using a Hamiltonian formulation and forward-backward sweep.

Referring to the models used for the diseases before SARS-CoV-2, recent research is done in order to understand more about the characteristics and behavior of the novel coronavirus. In [20], the authors propose to predict the new coronavirus infection based on a modified SEIR model since they assume that coronoavirus is infectious during the incubation period. They analyzed the situation in China and predicted that the number of infectious people will reach about $2.384.803$ people and that social isolation is very important to reduce as much as possible the epidemic size. At the end, it is predicted that this epidemic will last for 103 days and end on 13 March 2020. In [12], the authors analyze the case of SARS-CoV-2 for India till 30 March 2020 and predict the number of infected individuals for the next two weeks. SEIR model and Regression model have been used for forecasting. The authors evaluated the performance of both models using RMSLE and achieved 1.52 for the SEIR model and 1.75 for the regression model. It is found that R_0, which is the spread of the disease, was 2.02. Expected cases may rise between $5000 - 6000$ in the next two weeks of time and looking at the trend, there was definitely going to be an increase in the number of cases and the peak was yet to come after 30 March 2020. In [16], the authors fit a deterministic SEIR meta-population transmission model of infection within and between major Chinese cities to the daily number of confirmed cases of CoV-2019 in Chinese cities and cases reported in other countries/regions, using an assumption of Poisson-distributed daily time increments. The authors estimate a basic reproductive number of 3.11 and 58-76% of transmissions must be prevented to stop increase. The results show that the estimates of the basic reproductive number for this novel coronavirus to be comparable to most estimates reported for SARS and MERS-CoV, and also similar

to some estimates from subsets of data in the early period of SARS. At the end, it is found that the transmission rate in Wuhan was 1.94 while the infectious period was 1.61 days. They calculated the basic reproductive number of the infection to be 3.11 comparable to the range for SARS estimated from outbreaks during the 2003 epidemic (see [10]). In our paper, we propose to use a SIR model by applying six numerical techniques to solve ODEs [1] in order to modelize and study the behavior of SARS-CoV-2 in Nordic countries such as Norway, Sweden, Finland and Denmark. According to [15], SIR and SEIR models perform almost in the same way, i.e. they have same maximum number and a convergence of the curve of infectious at the same time. While [3] propose a time delay dynamical model for the outbreak of SARS-CoV-2 since according to them, this epidemic can be spread during the latent period and therefore classical models such as SIR, SEIR and SEIJR are not appropriate to describe the outbreak of SARS-CoV-2 in China. The objectives of this study are:

- Understanding the origin and diffusion of SARS-CoV-2 all over the world and when it is possible to indicate any connection with Nordic countries.
- Finding the rate of the spread of the disease in Nordic countries.
- Developing a mathematical SIR model to evaluate the spread of the disease.
- Prediction of the SARS-CoV-2 outbreak with SIR model by applying different numerical techniques to solve the ODEs.

After the introductory background in Section I, Section II presents the origin and the diffusion of the SARS-CoV-2 epidemic by analyzing the phylogenetic tree. A comparison of ODEs numerical solutions (theoretically) and methodology about the model used in this study is performed in Section III and Section IV. Section V covers analysis, experimental results and performance evaluation, and section VI concludes discussing our main findings.

II. The Origin and the Diffusion of the SARS-CoV-2 Epidemic

SARS (Severe Acute Respiratory Syndrome) is the recently-emerged disease caused by a new type of coronavirus (SARS-CoV-2). It consists of a 27 to 34 kilo bases-long, single-stranded RNA and displays a characteristic spiky envelope protein that resembles a crown. The first cases of SARS appeared in late 2019 in Wuhan, the capital of China's Hubei and grew into a major outbreak in the next few months (December, 2019 to 2020). The first reported infection has been unofficially reported to have occurred on 17 November 2019. The majority of the infected individuals acquired the disease in the Wuhan province of China. In early and mid-January 2020, the virus spread to other Chinese provinces, helped by the Chinese New Year migration and Wuhan being a transport hub and major rail interchange. From mid-January 2020, the epidemic started to spread outside China and become an international pandemic.

[1]Ordinary differential equations

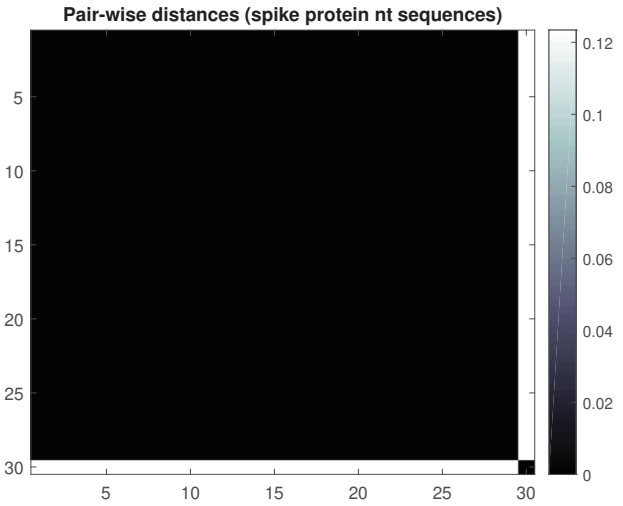

Fig. 4 The symmetric matrix that holds pair-wise genetic distances with Jukes-Cantor corrections. The dark color displays cases who are very similar to each other while the lighter one cases who are different from each other.

By analyzing the phylogenetic relationships between the samples of SARS-CoV-2 that were collected in late 2019 and in 2020, we can reconstruct the history of the SARS-CoV-2 epidemic and understand how it spread throughout the world in such a short period of time [4], [8].

A. The Sequence Pair-Wise Distances and Neighbor-Joining Phylogenetic Tree

We consider the nucleotide sequences of 447 strains of human SARS-CoV-2 for which the location and the date of collection are known. For practical illustration in our study, we have chosen only nucleotide sequences of 30 people including two cases from Nordic countries (Finland and Sweden). The sequences correspond to the spike S protein, which is responsible for binding to specific r eceptors a nd is considered a major antigenic determinant. Because the bat is believed to be the source of the human SARS-CoV-2, we also consider the sample derived from the bat [13]. This data are available in the website of National Center for Biotechnology Information (NCBI([21])). In order to obtain the distance matrix, we need to build the phylogenetic tree by computing a symmetric matrix that holds pair-wise genetic distances with Jukes-Cantor corrections ignoring sequence sites representing gaps. By plotting the distance matrix (see Fig. 4), we can appreciate the presence of a subset of sequences that are more closely related to each other (represented by the darker tones).

The last sequence, which is associated to the bat, is the most distant to all members of the set and this is expected because bat is a nonhuman coronavirus. Using the distances computed above, we construct a phylogenetic tree using the neighbor-joining method. In this case, we assume equal variance and independence of evolutionary distance estimates.

The tree (Fig. 5) describes the story of the epidemic. From the samples under study, the early infections occurred in

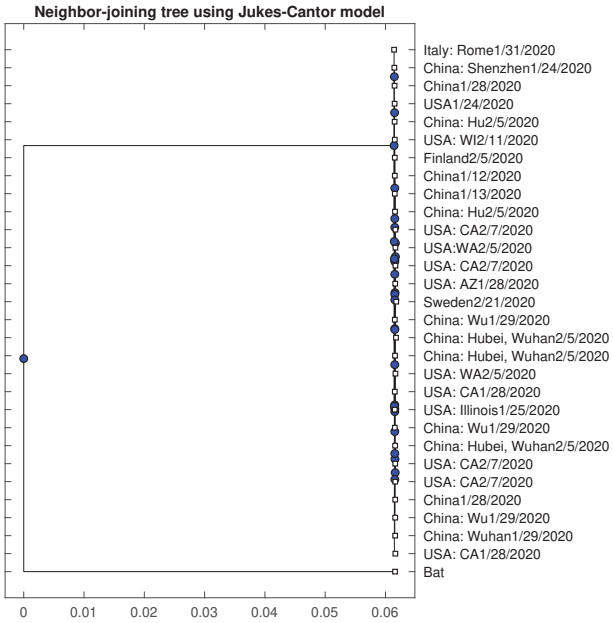

Fig. 5 The phylogenetic tree using the neighbor-joining method

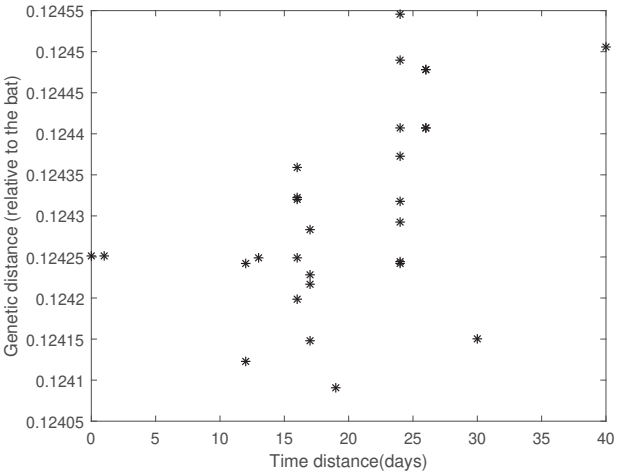

Fig. 6 The progression of the virus mutations over time

China, USA and Italy. All other cases are related to each other and most of them seem to branch from the cases traced back in China and Italy.

B. Estimating the Date of Origin of the Epidemic

Because the date of collection of each SARS-CoV-2 strain is known, we can observe the progression of the virus mutations over time. Consider the pair-wise distances according to the Kimura model, which distinguishes between transitional and tranversional mutation rates. After that, we restrict our analysis to the distance of each human strain from the bat strain. At the end, we plot the genetic distance versus the date of collection (Fig. 6).

In relation to the sequence of the bat, we observe that

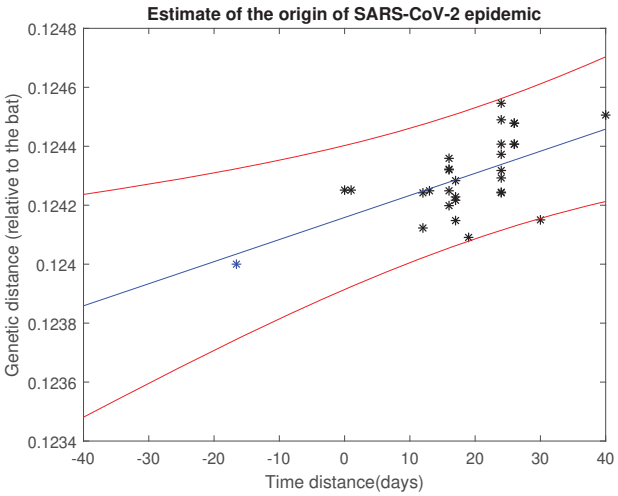

Fig. 7 Estimation of the origin of SARS-CoV-2 epidemic from polynomial fitting and a least-square interpolation

the genetic distance increases approximately in a linear manner with time. To outline the progression of the viral mutations over time and estimate the approximate date for the origin of the epidemic, we perform a polynomial fitting and a least-square interpolation. The start of the infection corresponds more or less to the root of the polynomial fit, i.e., any date that is at zero genetic distance from the bat's sequence (Fig. 7). The point in blue color indicates the estimated origin.

C. Rerooting the Phylogenetic Tree

Because the disease caused by the novel strain of human SARS-CoV-2 appears to have originated in the bat, we can assume that the location of the root for the human strains phylogenetic tree is next to the node associated with the bat. The rerooted tree (see Fig. 8) illustrates better the progression of the SARS-CoV-2 epidemic. Starting with the early infections in the Wuhan province of China, the virus spread in the Italy (01/31/2020) and USA (01/24/2020). From the USA (02/11/20202), the virus arrives to Sweden (02/21/2020), while from Italy (01/31/2020), the virus arrives to Finland (02/05/2020) and from Finland to the USA again. These results are and are not in line with official declaration of countries, for instance it is declared that coronavirus in the USA arrived from European countries, but as we see from the chart, some cases in the USA are related directly with bat genomes, while from the other side, there are other cases that are branches of European countries. Note that all these results come from the chosen dataset of 30 cases from 447 available set of nucleotide sequences in NCBI.

III. Visualizing the Diffusion of the Virus via a Directed Graph

We can also visualize the diffusion of the virus using a directed graph, where each node represents an infected individual and weights of edges are associated to the genetic distance between sequences. First, create an adjacency matrix

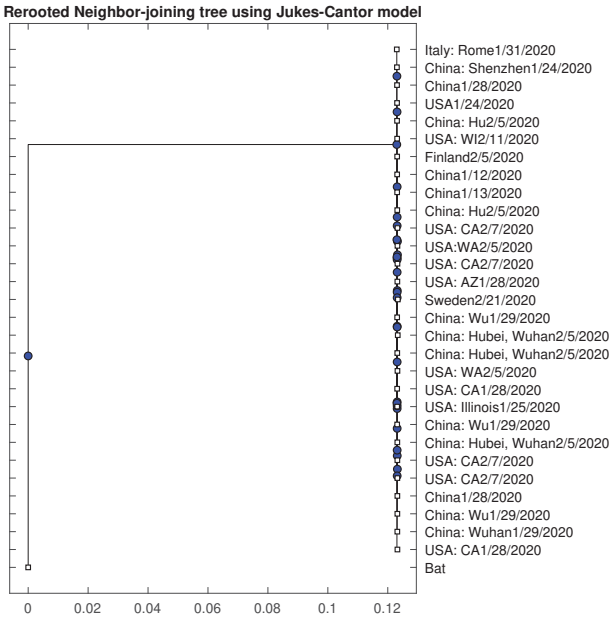

Fig. 8 Rerooting the phylogenetic tree of the progression of the SARS-CoV-2 epidemic

based on the date of collection of the samples, such that possible paths run through nodes that are compatible in terms of the collection dates. Then, use the previously computed Jukes-Cantor distances to assign weights to the edges between nodes. And finally, determine the shortest path from the node associated with the bat and every other node.

This graph highlights the crucial role played by some of the infection events:

- The bat is the source of the infection.
- China and Italy were identified as roots of the branching for most of the international epidemic.
- Italy and the USA represent the bridge connecting the spread of the virus in Nordic countries.

IV. A Comparison of ODEs Numerical Solutions

Problems involving ODEs can always be reduced to the study of sets of first-order differential equations. For example, the second-order equation:

$$\frac{d^2y}{dt^2} + q(x)\frac{dy}{dt} = r(x) \tag{1}$$

can be rewritten as two first-order equations,

$$\frac{dy}{dt} = z(t) \tag{2}$$

$$\frac{dz}{dt} = r(t) - q(t)z(t) \tag{3}$$

where z is a new variable. This exemplifies the procedure for an arbitrary ODE. The usual choice for the new variables is to let them be just derivatives of each other (and of the original variable). The generic problem in ordinary differential equations is thus reduced to the study of a set of N coupled first-order differential equations for the functions $y_i, i = 1, 2, ..., N$, having the general form:

$$\frac{dy_i(t)}{dt} = f_i(t, y_1, ..., y_N), i = 1, ..., N \tag{4}$$

where the functions f_i are known. Given the equations and the initial point, the basic idea for estimating the system is to rewrite the equations in terms of finite s teps. T he initial conditions are used to extrapolate the solution to the next step. In the limit, as the step size approaches zero a good approximation comes. Euler's method is a first o rder method that follows this approach; however, aside from demonstrating the concept, Eurler's method is not useful as a serious numerical technique. Runge-Kutta and predictor-corrector methods are the most common general methods [11].

Runge-Kutta methods sample the solution at steps smaller than the overall step size. These samples are then included in the estimate when the entire step is extrapolated. The sub-step samples are used to match the coefficients of a higher-order Taylor series expansion. The fourth order Runge-Kutta method is often referred to as RK4 which uses four samples to calculate a solution for the value at $t(n+1)$ one sample at $t(n)$ two samples at $t(n+1/2)$ and one sample at $t(n+1)$. These sample values are then used in a polynomial equation that extrapolates the system from $t(n)$ to $t(n+1)$. In most applications, RK4 will be able to solve the system. This makes RK4 the method of choice for problems where computational efficiency is not a big concern. Higher order Runge-Kutta methods exist but it is more common to use a predictor-corrector method instead of higher order Runge-Kutta methods. The predictor-corrector methods use a predictor phase to make the problem appear as if we are integrating a function. The corrector phase uses this predicted value to refine the answer. Predictor-corrector methods use one polynomial to extrapolate the solution to the next point and another polynomial to add the correction, and this correction step can be run twice or even iteratively. For reasonably smooth problems, predictor-corrector methods are more efficient compared to Runge-Kutta methods. The most popular implementation for predictor-corrector methods is the Adams-Bashforth-Moulton method where Adams-Bashforth is the predictor and Adams-Moulton is the corrector.

The formula for Euler method is:

$$y(n+1) = y(n) + hf(t(n), y(n)) \tag{5}$$

which advances a solution $t(n)$ to $t(n+1) = t(n) + h$. The solution is advanced through an interval h, but uses derivative information only at the beginning of that interval. It means that the step error is only one power of h smaller than the correction, i.e $O(h^2)$ added to (5). Euler's method is not recommended for practical use, because the method is not very accurate when compared to others, and neither is it very stable. Consider, however, the use of a step like (5) to take a "trial" step to the midpoint of the interval. Then use the value of both t and y at that midpoint to compute the "real" step across the

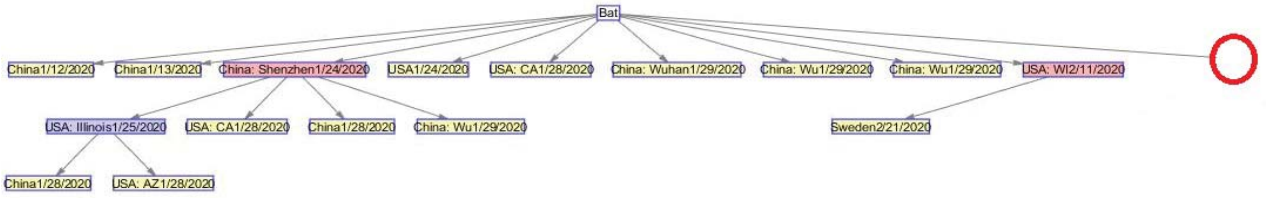

Fig. 9 The diffusion of the virus via a directed graph using Jukes-Cantor distances: part 1

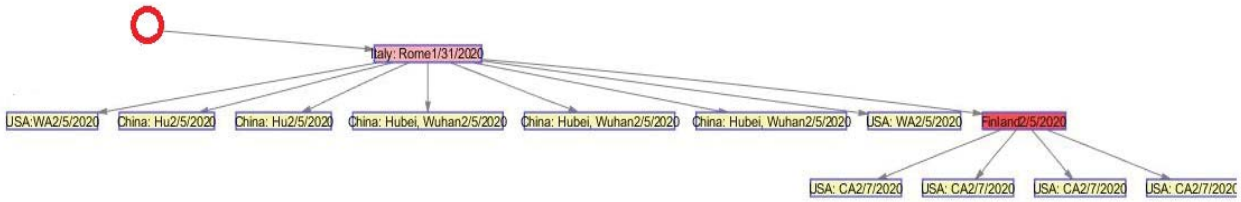

Fig. 10 The diffusion of the virus via a directed graph using Jukes-Cantor distances: part 2

whole interval. In equations, it can be expressed as follows:

$$k_1 = hf(t(n), y(n)) \tag{6}$$

$$k_2 = hf(t(n) + \frac{1}{2}h, y(n) + \frac{1}{2}k_1) \tag{7}$$

$$y(n+1) = y(n) + k_2 + O(h^3) \tag{8}$$

As indicated in the error term, this symmetrization cancels out the first-order error term, making the method second order, and we know that a method is conventionally called n*th* order if its error term is $O(h^{n+1})$. In fact, the system of (2)-(4), is called the second-order Runge-Kutta or midpoint method. There are many ways to evaluate the right-hand side $f(t, y)$, which all agree to first order, but that have different coefficients of higher-order error terms. Adding up the right combination of these, we can eliminate the error terms order by order. That is the basic idea of the Runge-Kutta method. But the most-often used method is the classical *fourth-order Runge-Kutta method (RK4)*. The system of equations for this method are given as follows:

$$k_1 = hf(t(n), y(n)) \tag{9}$$

$$k_2 = hf(t(n) + \frac{h}{2}, y(n) + \frac{k_1}{2}) \tag{10}$$

$$k_3 = hf(t(n) + \frac{h}{2}, y(n) + \frac{k_2}{2}) \tag{11}$$

$$k_4 = hf(t(n) + h, y(n) + k_3) \tag{12}$$

$$y(n+1) = y(n) + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6} + O(h^5) \tag{13}$$

The RK4 method requires four evaluations of the right hand side per step h. This will be superior to the midpoint method (8) if at least twice as large a step is possible with (13) for the same accuracy. In [5], the authors introduce an improvement of the RK4 method, adding some more accuracy in this method. More specifically, it uses six function evaluations to calculate fourth and fifth order accurate solutions. The difference between these solutions is then taken to be the error of the fourth order solution. Dormand and Prince chose the coefficients of their method to minimize the error of the fifth order solution. The one step calculation in *the Dormand-Prince (RK(4,5)) method* is done as follows:

$$\begin{aligned}
k_1 &= hf(t(n), y(n)) \\
k_2 &= hf(t(n) + \frac{h}{5}, y(n) + \frac{k_1}{5}) \\
k_3 &= hf(t(n) + \frac{3h}{10}, y(n) + \frac{3k_1}{40} + \frac{9k_2}{40}) \\
k_4 &= hf(t(n) + \frac{4h}{5}, y(n) + \frac{44k_1}{45} - \frac{56k_2}{15} + \frac{32k_3}{9}) \\
k_5 &= hf(t(n) + \frac{8h}{9}, y(n) + \frac{19372k_1}{6561} - \frac{25360k_2}{2187} \\
&\quad + \frac{64448k_3}{6561} - \frac{212k_4}{729}) \\
k_6 &= hf(t(n) + h, y(n) + \frac{9017k_1}{3168} - \frac{355k_2}{33} - \frac{46732k_3}{5247} \\
&\quad + \frac{49k_4}{176} - \frac{5103k_4}{18656}) \\
k_7 &= hf(t(n) + h, y(n) + \frac{35k_1}{384} + \frac{500k_3}{1113} \\
&\quad + \frac{125k_4}{192} - \frac{2187k_5}{6784} - \frac{11k_6}{84})
\end{aligned} \tag{14}$$

Then the value $y(n+1)$ is calculated as:

$$\begin{aligned}
y(n+1) = y(n) + \frac{35k_1}{384} + \frac{500k_3}{1113} + \frac{125k_4}{192} \\
- \frac{2187k_5}{6784} + \frac{11k_6}{84}
\end{aligned} \tag{15}$$

This is the calculation by RK4; next we will calculate the value z(n+1) by RK method of order 5 as:

$$z(n+1) = y(n) + \frac{5179k_1}{57600} + \frac{7571k_3}{16695} + \frac{393k_4}{640} - \frac{92097k_5}{339200} + \frac{187k_6}{2100} + \frac{k_7}{40} \quad (16)$$

The difference $error = |z(n+1) - y(n+1)|$ is calculated:

$$error = |\frac{71k_1}{57600} - 71k_3 16695 + \frac{71k_4}{1920} - \frac{17253k_5}{339200} + \frac{22k_6}{525} - \frac{k_7}{40}| \quad (17)$$

This is considered as error in $y(n+1)$, the optimal time interval in the next step of the calculation h_{opt} will be expressed as $h_{opt} = s*h$ where $s = (\frac{\epsilon h}{2*error})^{\frac{1}{5}}$. Note that high order does not always mean high accuracy. Predictor-corrector methods can be much more efficient for problems where very high accuracy is a requirement.

A predictor-corrector method such as *Adams-Bashforth-Moulton (ABM) method* is used to solve an ODE in two steps [18]. First, the prediction step calculates a rough approximation of the desired quantity using an explicit method, the Adams-Bashforth method. Second, the corrector step refines t he i nitial a pproximation u sing an implicit method, the Adams-Moulton method. Let us define the two step Adams method of order p.

The Adams-Bashforth formula of order p is obtained by integrating the polynomial $P(t)$ that interpolates $f(n+1-i)$ at $t(n+1-i)$ for $i = 1, 2, ..., p$ in place of f:

$$y(n+1) = y(n) + h\sum_{i=1}^{p} \alpha(p,i) f(n+1-i) \quad (18)$$

The implicit Adams-Moulton formula arises when the polynomial $P(t)$ interpolates $f(n+1-i)$ for $i = 0, 1, ..., p-1$

$$y(n+1) = y(n) + h\sum_{i=1}^{p-1} \beta(p,i) f(n+1-i) \quad (19)$$

When $i = p-1$ the right hand side contains the term $f(n+1) = f(t(n+1), y(n+1))$ and it is clear that $y(n+1)$ is defined only implicitly by this formula. The solution is given by first predicting the result using the explicit Adams-Bashforth formula and then correcting it using the implicit Adams-Moulton formula.

In order to solve stiff problems of the form:

$$M(t)y' = f(t,y) \quad (20)$$

the Klopfenstein-Shampine (KSH) and the modified Rosenbrock triple (MRT) methods are used, where $M(t)$ is the non-singular matrix mass and usually sparse. The *KSH method* is a quasi-constant step size implementation of the numerical differentiation formulas (NDF) in terms of backward differences [18], [19]. Options allow integration with backward differences formulas (BDF) and integration with a maximum order less than five. K lopfenstein s tudied methods of the following form that he called NDF:

$$\sum_{m=1}^{k} \frac{1}{m} \nabla^m y(n+1) = -hf(t(n+1), y(n+1)) - \kappa\gamma(k)(y(n+1) - y^{(0)}(n+1)) \quad (21)$$

where k is a scalar parameter and the coefficients $\gamma(k)$ are given by:

$$\gamma(k) = \sum_{j=1}^{k} \frac{1}{j}$$

The role of the term added to BDF of order k is illuminated by the identity:

$$y(n+1) - y^{(0)}(n+1) = \nabla^{k+1} y(n+1) \quad (22)$$

The *MRT method* is based in modified Rosenbrock formula of order 2. Since it is a single-step solver, this method may be more efficient than the KSH method at solving problems with solutions that change rapidly and it can solve some kinds of stiff problems for which KSH is not effective. The MRT method evaluates the Jacobian during each step of the integration, so supplying it with the Jacobian matrix is critical to its reliability and efficiency. The MRT formula is expressed in the following form:

$$F(0) = F(t(n), y(n)) \quad (23)$$

$$k_1 = W^(-1)(F(0) + hdT) \quad (24)$$

$$F(1) = F(t(n) + 0.5h, y(n) + 0.5hk_1) \quad (25)$$

$$k_2 = W^(-1)(F(1) - k_1) + k_1 \quad (26)$$

$$y(n+1) = y(n) + hk_2 \quad (27)$$

$$F(2) = F(t(n+1), y(n+1)) \quad (28)$$

$$k_3 = W^{-1}[F(2) - e_{32}(k_2 - F(1)) - 2(k_1 - F(0)) + hdT] \quad (29)$$

where $W = I - hdJ$ with $d = 1/(2+\sqrt{2})$ and $e_{32} = 6 + \sqrt{2}$, $J \approx \frac{\delta F}{\delta y}(t(n), y(n))$, $T \approx \frac{\delta F}{\delta t}(t(n), y(n))$ and h is a constant step size.

The *trapezoidal rule (TR) method* is based on a variant of the backward differentiation formulas, BDFs, called numerical differentiation formulas, NDFs. This method is an implementation of the trapezoidal rule using a "free" interpolant. We can use this method in cases when the problem is moderately stiff and it is needed a solution without numerical damping. It was developed to solve the DAEs and integrate stiff ODEs of the form:

$$M(t)y' = f(t,y) \quad (30)$$

where the mass matrix $M(t)$ is singular. Solving a DAE is more complicated than solving an ODE because a DAE has

a solution only if the initial conditions $y(0)$ are consistent in the sense that the equation $M(t(0))y'(0) = f(t(0), y(0))$ has a solution $y'(0)$ for the initial slope [19].

The implicit multi-step *trapezoidal rule - backward differential formula of order 2 (TR-BDF2)* is developed for the simulation of circuits and semiconductor devices. The method is an implicit Runge-Kutta formula with a first s tage t hat is a trapezoidal rule step and a second stage that is a backward differentiation formula of second order. It is applicable for stiff problems of low accuracy because it is not strongly stable. Referring to [6] and [1], the TR-BDF2 method is expressed as follows:

$$y(n+\gamma) = y(n) + \gamma h[(1-\theta)f(t(n), y(n) + \theta f(t(n+\gamma), y(n+\gamma))], \quad 0 < \theta \leq 1$$

This method is the only method from the Runge-Kutta methods family which admits an embedded, asymptotically correct error estimate.

After review of numerical solutions, we arrive at the conclusion that, Runge-Kutta methods of order less then four are not preferred for use in differential equations problems [11]; they can be used just as an introduction with numerical methods, since they display a poor error performance. Better error performance would be preferred and for this, RK(4,5) is the one of the first methods which supports adaptive steps size. According to [5], RK(4,5) is the best method to apply as a first try for many problems. It is a one-step method and in order to compute $y(n)$, it needs only the solution at the immediately preceding time point $y(n-1)$. ABM is preferred over RK(4,5) because it may be more efficient at stringent tolerances than RK(4,5) when low error and computational efficiency are desired. Higher-order ABM generally lowers errors even further. In contrast to RK(4,5), the ABM method is a multi-step method, and in order to compute the current solution it needs the solutions at several preceding time points. These two methods are intended to solve non-stiff systems. If RK(4,5) and ABM fail to converge or runs too slowly then one can use the KSH method which is a variable order method based on the NDFs. Like the ABM method, the KSH method is a multi-step method. In cases when RK(4,5) fails or is very inefficient and there is no doubt that the problem is stiff, one may use the KSH method. MRT has advantages when one or more of the system eigenvalues are near the jw axis. Because it is a one-step method, it may be more efficient than KSH at crude tolerances. It can solve some kinds of stiff problems, for which KSH is not effective. TR method is an implementation of the trapezoidal rule using a free interpolant. This method can be used if the problem is only moderately stiff and it is needed a solution without numerical damping. Bulirsch-Stoer is another numerical solution of ODEs which is identified as possibly a better alternative to the ABM method and should be suitable when high accuracy is desired [14]; however, we have not considered it in our study.

V. Simulation of the Epidemic Outbreak Based on SIR Model

A susceptible-infected-recovered (SIR) model is used to simulate an epidemic breakout. In order to comprehend the disease transmission, we decided to start our study about coronavirus based on a SIR model without vital dynamics which is a dynamical system that separate a population in three sub-clusters. The variables that are considered in this model are:

- Susceptible cases S(t)
- Infectious cases I(t)
- Recovered cases R(t)

while the parameters are as follows:

- β – infection/ transmission rate
- γ – inverse of the average latent time

In the deterministic form, the SIR model can be written as the following ODEs:

$$\frac{dS}{dt} = -\frac{\beta S(t) I(t)}{N} \tag{31}$$

$$\frac{dI}{dt} = \frac{\beta S(t) I(t)}{N} - \gamma I(t) \tag{32}$$

$$\frac{dR}{dt} = \gamma I(t) \tag{33}$$

where $N = S(t) + I(t) + R(t)$ is the total population and it is assumed constant because if we get the first derivative of the sum of the infected, susceptible and recovered dynamics, it will be equal to zero:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = -\frac{\beta S(t) I(t)}{N} + \frac{\beta S(t) I(t)}{N} - \gamma I(t) + \gamma I(t) = 0$$

We do not have to worry about the fact that N is an assumed constant since the growth rate of the population compared to the epidemic duration is very low. An epidemic is just for a short period of time for instance one month, two months or a season while the population grows over years. The infectious rate, β, controls the rate of spread which represents the probability of transmitting disease between a susceptible and an infectious individual. The dynamics of the infectious people depends on the basic reproduction ratio:

$$R_0 = \frac{\beta}{\gamma}$$

It shows the average number of contacts by an infected person with others before the infected person moves to the recovered status. At the beginning of the epidemic, there are no recovered people, thus the initial conditions will be written as follows:

$$(S(0) = S_0, I(0) = I_0, R(0) = 0) \Rightarrow$$
$$N = S(0) + R(0) + I(0) = S_0 + I_0$$

Moreover, at the beginning of the epidemic, it will be just a small number of infected people, or even just one, so the initial number of infected people will be relatively a small number, while the number of susceptible people in the beginning will be a large number. Therefore, from the system of differential equations, we understand that at the very beginning of the epidemic we have susceptible people which become infected,

and infected people who create more infected cases where some of those infected people become recovered. So, the more people are connected together, the more the susceptible and infected people are going to multiply together to create more interactions, and as a result, the higher the likelihood that susceptible people will become infected. From (37), the rate of changes of susceptible people is negative because the susceptible cases become smaller with time and they became infected cases. The more infected there are, the higher the rate of change of infectious leaving. At the end, infectious people turn to recovered people. To understand more about the epidemic spread and size, let us consider the dynamics of infected people at time $t = 0$.

$$\frac{dI}{dt}|_{t=0} = \frac{\beta S_0 I_0}{N} - \gamma I_0$$

If $\frac{dI}{dt}|_{t=0} > 0$, then the disease will result in an epidemic, otherwise new people may get infected but they leave the status of being infected fast before they become infectious. This is equivalent with the inequality $\frac{\beta \mathbf{S_0}}{\gamma} > \mathbf{1}$. In order to reduce the epidemic size, we need to lower the transmission rate β and this is done by prevented measures undertaken. Another way is to lower the initial susceptible people by vaccinating people. Let us consider again the dynamic changes of infected people but assuming that susceptible people remain approximately constant:

$$\frac{dI}{dt}|_{t=0} = \frac{\beta S_0 I(t)}{N} - \gamma I(t)$$

A possible solution of this differential equation is:

$$I(t) = \exp^{(\beta S_0 - \gamma)t}$$

This means that even in the beginning when $t = 0$ and the susceptible people is approximately constant, the epidemic has an exponential growth.

Now, let us consider (37) and (38) in order to observe the maximum number of infected at any given time. For this, we divide (38) by (37):

$$\frac{dI}{dS} = \frac{\beta IS - \gamma I}{-\beta IS} = -1 + \frac{\gamma}{\beta S} = -1 + \frac{1}{R_0 S}$$

After integration we obtain:

$$I + S - \frac{1}{R_0} lnS = I_0 + S_0 - \frac{1}{R_0} lnS_0 \qquad (34)$$

To find the maximum number of infected people from the epidemic, we need to differentiate the function that indicates the number of infected people and equal it to zero.

$$\frac{dI}{dS} = -1 + \frac{1}{R_0 S} = 0 \Rightarrow S = \frac{1}{R_0}$$

Therefore, the maximum number of infected will come by solving (34) substituting $S = {}^{1}$: $\overline{R_0}$

$$I_{max} = I_0 + S_0 - \frac{1}{R_0}(1 + ln(R_0 S_0))$$

The maximum number of infected people at a given time is equal to the total population $(I_0 + S_0)$ subtracting a function of R_0. Regarding the total number of susceptible people left at the end of the epidemic, we are going to reconsider the fact that we are assuming a constant population. So, $R + I + S = I_0 + S_0$, which implies $R(end) = I_0 + S_0 - S(end) - I(end)$. At the end of the epidemic, we assume that the number of infected people will go to zero, so $R(end) = I_0 + S_0 - S(end)$ and now we focus to calculate the value of $S(end)$. For this, we go back to (34), substituting S with $S(end)$, we find out:

$$S(end) - \frac{1}{R_0} ln(S(end)) = I_0 + S_0 - \frac{1}{R_0}(1 + ln(S_0))$$

At the end, after we find the total of susceptible number, we can calculate the total number of people who were infected by the disease as $R(end) = I_0 + S_0 - S(end)$.

The system of ODEs in matrix form can be expressed as follows:

$$\frac{dY}{dt} = A * Y + B$$

where $Y = [S, I, R]^T$

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\gamma & 0 \\ 0 & \gamma & 0 \end{bmatrix} \quad B = S(t)I(t) \begin{bmatrix} -\frac{\beta}{N} \\ \frac{\beta}{N} \\ 0 \end{bmatrix}$$

The equation $\frac{dY}{dt} = A * Y + B$ is then solved using the classic six different ODE numerical methods such as RK(4,5), ABM, KSH, MRT, TR and TR-BDF2. In the next section, one will find an analysis of the epidemic outbreak comparing ODE numerical methods and predicting the future of this epidemic/pandemic.

VI. Forecasting of SARS-CoV-2 Epidemic with SIR Model

For the estimation of the epidemic evolution, we are using the SIR epidemic model presented in the previous section. The model assumes a constant population and uniform mixing of the people. The model is data-driven, so its forecast is as good as the source of the data and the results of the model will be improved day by day with new data added in the dataset. Our data are confirmed, deaths, and recovered cases per day published in WHO ([22]) from 22 January 2020 to 21 April 2020. The parameters of the model are obtained by minimization of the objective function using the Nelder-Mead simplex algorithm. On the epidemic evolution graph, we display some colored regions which separate the epidemic phases. It starts with the fast growth phase which is identified i n r ed c olor, t he t ransition t o s table-state phase identified i n y ellow c olor, t he e nding p hase i n g reen a nd the transition phase when the coronavirus could continue/restart the activity in orange.

Tables I-VI display results related to Nordic countries, China and Italy for all numerical methods under study. N indicates the initial size of population that has been susceptible of being positive with coronavirus, R_0 is the basic reproduction ratio which shows the expected number of new infections from a single infection in a population where all subjects are susceptible. These new infections are called secondary infections. CP is time between contacts, IP is

typical time until recovery, and $RMSE$ is the root mean square error. Note that all these variables are method-driven so the forecasts are as good as the accuracy of method used. Therefore, the end of an active cycle of the coronavirus epidemic is indicated by the decline in the number of infected individuals rather than an absolute lack of susceptible cases. Through numbers in the tables are reflected the measures the governments took during the development of the epidemic but also, we need to emphasize that the definition of an epidemic size depends on the population of a country N as well. In addition, between outputs, you will find that the susceptible cases follow an exponential distribution while unsusceptible cases follow a logarithmic distribution for all countries. The proportions of susceptible and recovered individuals could be not zero due to the fact that at the end of the course of the epidemic not all individuals of the population will be recovered. Let us start our analysis in two countries, China and Italy, where the virus originated and where it spread in the very beginning of this pandemic. From Figs. 11 - 14, one can see that the outbreak in China started on 25 January 2020 (this is because we consider data from 22 January 2020) and on 31 January 2020, the outbreak started to accelerate arriving in an inflection point on 9 February 2020, on 19 February 2020 a stable growth starts and finally the ending phase comes on 9 March 2020. But, if we focus on figures produced by ABM method, we see that the virus in China is still active, although at low rates and the third colored region is in orange which emphasizes that the virus is still active.

Therefore, the ABM method traces the activity of the virus better (for the period of time later than the ending phase) than other numerical methods which predicts that the virus will cease its activity after 9 March 2020.

After predictions about the continuation of the infectious curve, we tried to do the same for the deaths and recoveries curves (for more see Fig. 15). It is predicted that on 28 March 2020, China had 3881 deaths and 68965 recovered individuals.

Fig. 16, shows the transition rate from the compartment of susceptible individuals to the compartment of infectious individuals which passed the peak for China and now it is almost in zero for the new infected cases per day and is stable for the accumulated infected cases day after day. While Fig. 17 indicates that the susceptible cases now are at zero and the unsusceptible cases are more than 250000 cases. In Fig. 18, one can find a summary about the infected, dead, and recovered cases in order to understand better the active cases which are shown in red and correspond to the number of infected cases removing dead and recovered cases. On 21 April 2020, the predicted active cases for China are at low rates.

Table I shows a summary of the most important epidemic indicators for all numerical solution considered in our study for China. From the table, we see that the technique with smaller RMSE is the ABM method. The basic reproduction ratio varies from 1.05 to 1.64, which means that the secondary infections from one infected person are around one to two people. While the typical time between contacts varies from 0.27 to 1.56 days and the average infectious period varies from 0.28 to 2.51 days. According to the RMSE for China, the numerical methods are listed (from the smallest to the highest value) as follows:

- ABM
- MRT
- RK(4,5)
- KSH
- TR-BDF2
- TR

TABLE I
EPIDEMIC INDICATORS FOR CHINA

Method	N	R_0	CP	IP	RMSE
RK(4,5)	753960	1.056	0.27	0.28	1976.38
ABM	128906	1.579	1.56	2.47	1908.98
KSH	123738	1.633	1.53	2.5	2091.18
MRT	490753	1.091	0.43	0.47	1918.28
TR	122934	1.641	1.53	2.51	2120.13
TR-BDF2	125590	1.614	1.5	2.43	2107.11

Since in the case of China, the ABM method resulted as the method which represents better the current situation, we start to interpret the results for Italy referring to the ABM method. From the chart produced by the ABM method, (with a closer inspection) one can see that the outbreak in Italy started on 22 February 2020 and on 18 March 2020: the outbreak started to accelerate arriving in an inflection point on 27 March 2020, and on 5 April 2020 a stable growth starts and finally the ending phase comes on 22 April 2020. But from the charts, the ABM method is not the best fit in the real data and therefore we need to interpret the results of the MRT method referring to the literature review when we require accuracy in prediction. From the chart produced by the MRT method, with a closer inspection one can see that the outbreak in Italy started on 22 February 2020 and on 14 March 2020 the outbreak started to accelerate arriving in an inflection point on 29 March 2020, and on 16 April 2020, a stable growth starts and finally the ending phase comes on 20 May 2020. The difference from method to method is almost one month.

From Fig. 23, it is predicted that on 29 June 2020, Italy will have 19015 deaths and 97670 recovered individuals.

Fig. 24 displays the transition rate from the compartment of susceptible individuals to the compartment of infectious individuals which passed the peak for Italy and now is in low rates less of than 500 cases for new infected cases per day and is stable for the accumulated infected cases day after day. While Fig. 25 indicates that the susceptible cases now are at zero and the unsusceptible cases are just a few cases. Fig. 26 shows a summary about the infected, dead, and recovered cases in order to better understand the active cases which are shown in red and corresponds to the number of infected cases removing dead and recovered cases. On 21 April 2020, the predicted active cases for Italy are at low rates.

Table II shows a summary of the most important epidemic indicators for all numerical solutions considered in the study for Italy. From the table we see that the technique with smaller RMSE is the MRT method. The basic reproduction ratio varies from 1.057 to 2.13, which means that the secondary infections from one infected person are around one to two people. While the typical time between contacts varies from 0.44 to 3.46

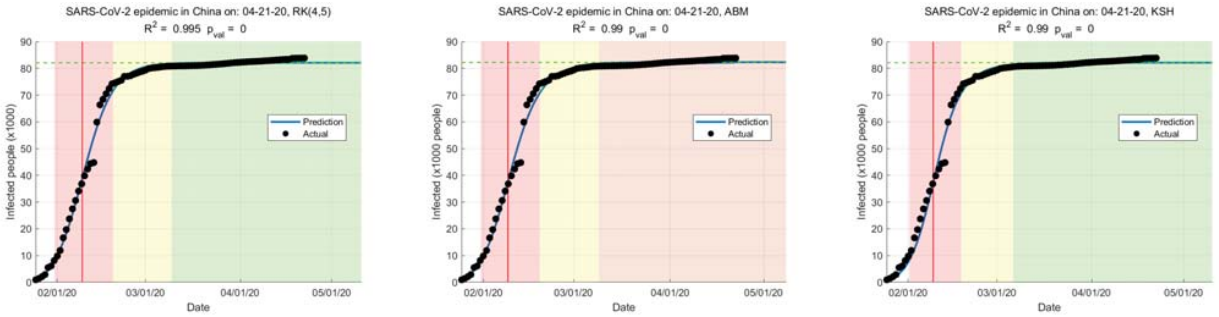

Fig. 11 Forecasting of SARS-CoV-2 epidemic by SIR model without vital dynamics for China

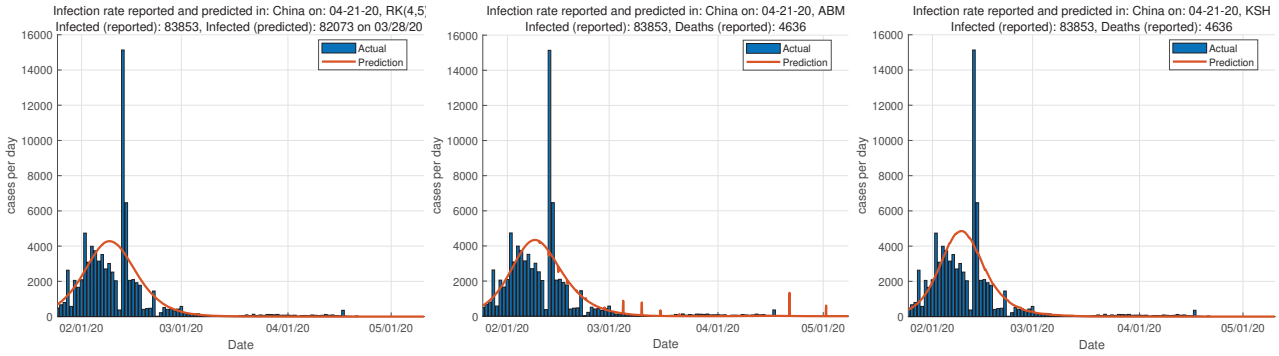

Fig. 12 The infection activity of SARS-CoV-2 predicted with SIR model without vital dynamics for China

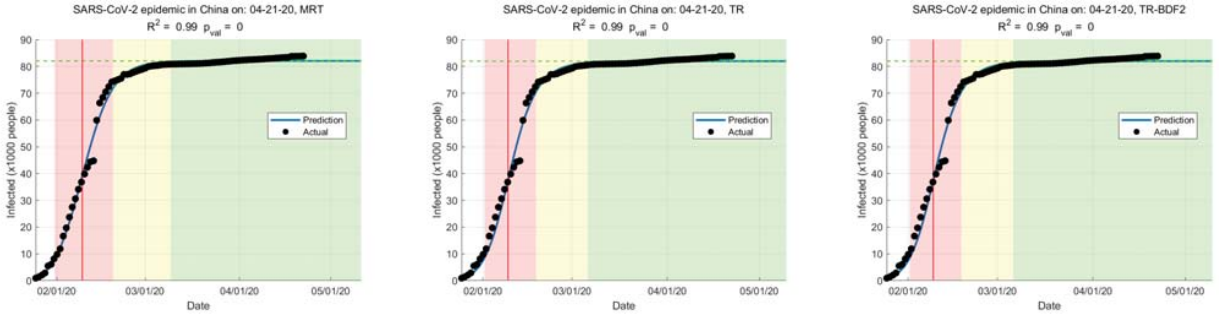

Fig. 13 Forecasting of SARS-CoV-2 epidemic by SIR model without vital dynamics for China

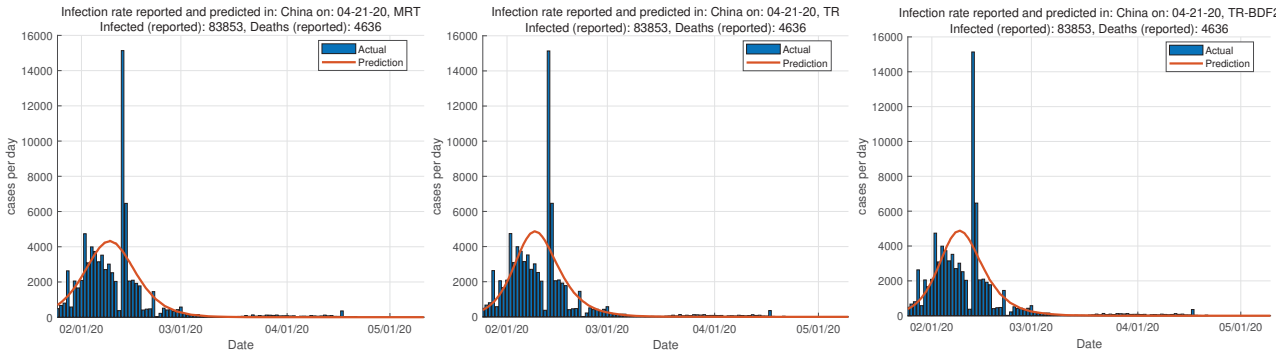

Fig. 14 The infection activity of SARS-CoV-2 predicted with SIR model without vital dynamics for China

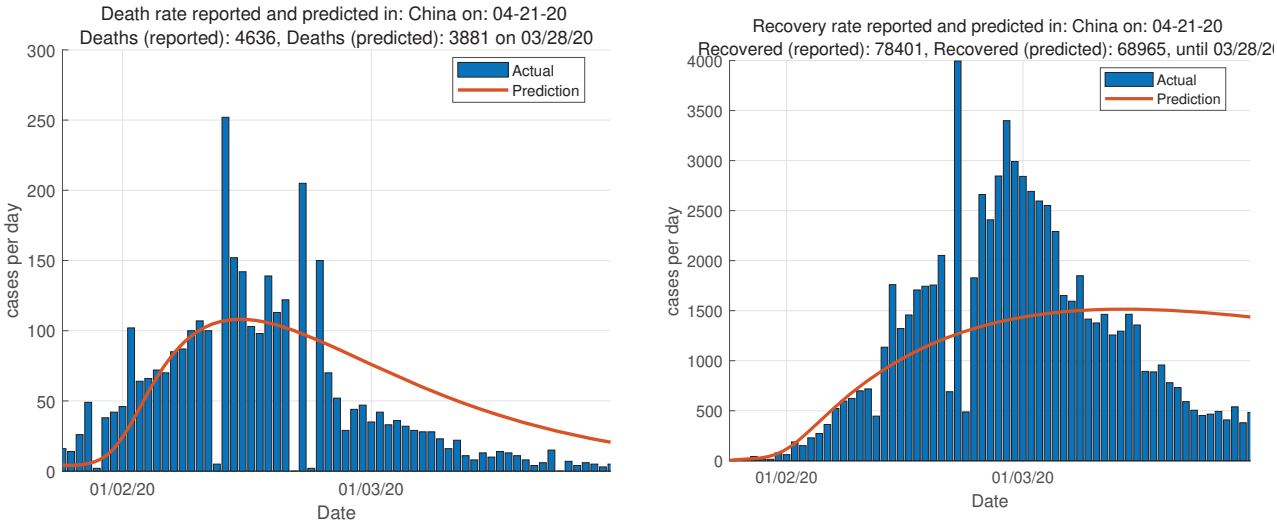

Fig. 15 Death and Recovery rate reported and predicted for China

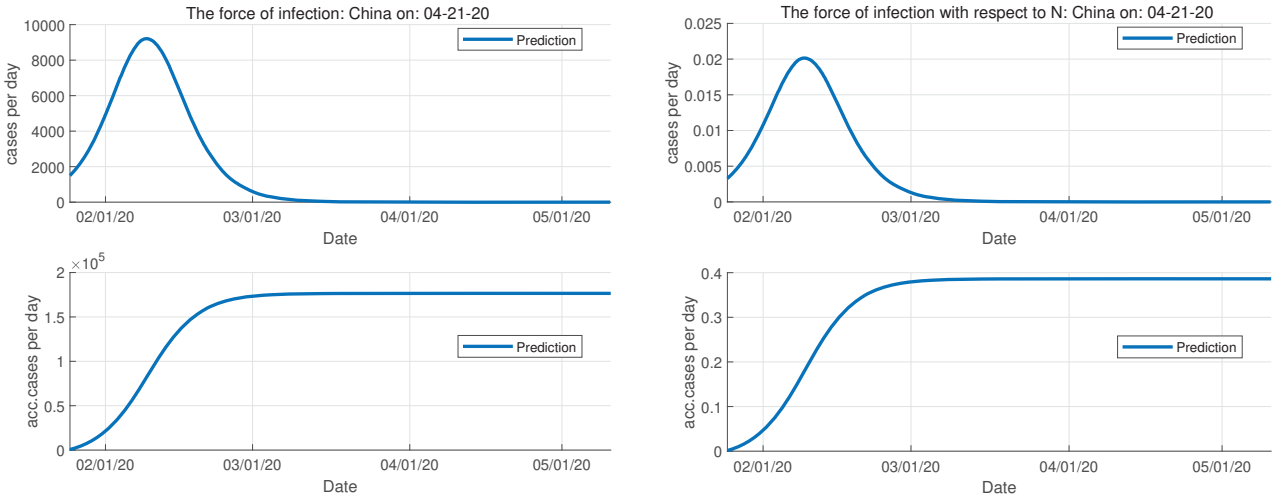

Fig. 16 The force of Infection which depends on the absolute number of infectious cases and on their fraction with respect to the total constant population for China

days and the average infectious period varies from 0.46 to 6.56 days. According to the RMSE for Italy, the numerical methods are sorted (from the smallest to the highest value) as follows:

- MRT
- RK(4,5)
- TR
- ABM
- KSH
- TR-BDF2

From the analysis done related to China and Italy and observing the outputs regarding to Denmark, it is interesting to comment on the results from the ABM and MRT methods (other methods reflect almost the same predictions). From the chart produced by the ABM method, (with a closer inspection) one can see that the outbreak in Denmark started on 6 March 2020 and on 24 March 2020 the outbreak started to accelerate

TABLE II
EPIDEMIC INDICATORS FOR ITALY

Method	N	R_0	CP	IP	RMSE
RK(4,5)	339936	1.459	2.34	3.41	3137.72
ABM	1737400	1.057	0.44	0.46	9908.22
KSH	333376	1.365	1.07	1.46	10933.1
MRT	251196	1.893	3.46	6.56	3083.58
TR	198256	2.137	2.3	4.92	8830.59
TR-BDF2	314057	1.395	1.11	1.55	11084.4

arriving at an inflection point on 3 April 2020, and on 14 April 2020 when a stable growth starts and finally the ending phase comes on 5 May 2020. But from the charts, this is not the best fit in the real data, therefore we need to interpret the results of the MRT method, since the smallest value of RMSE is achieved from this method. From the chart produced by the MRT method, with a closer inspection one can see that the outbreak in Denmark started on 6 March 2020 and on 20

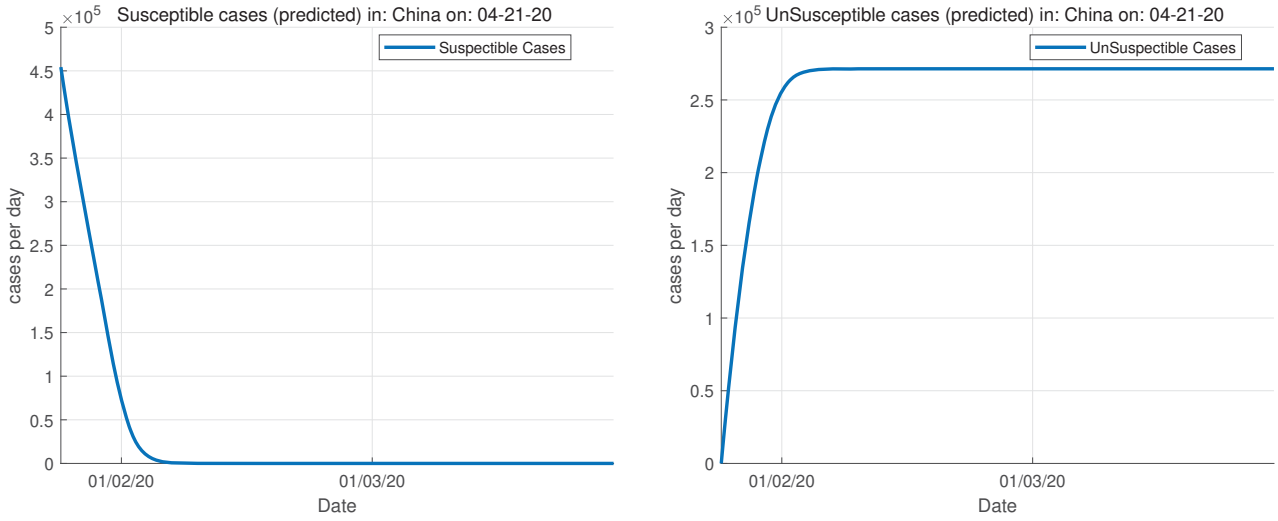

Fig. 17 All output channels for SIR model without vital dynamics for China

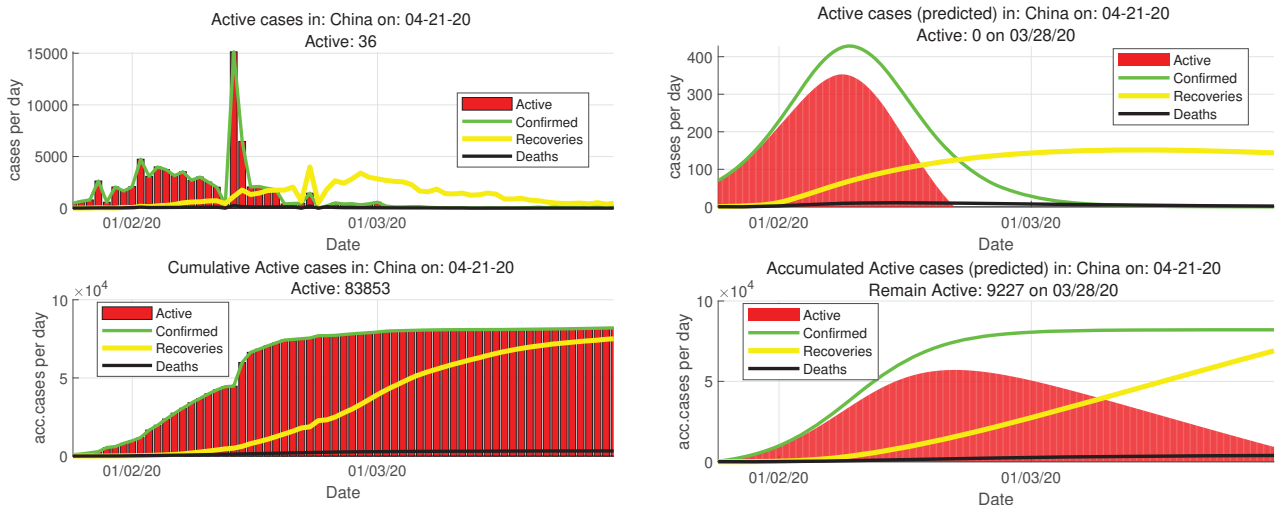

Fig. 18 Active cases reported and predicted from the SIR model without vital dynamics until the end of the activity of coronavirus for China

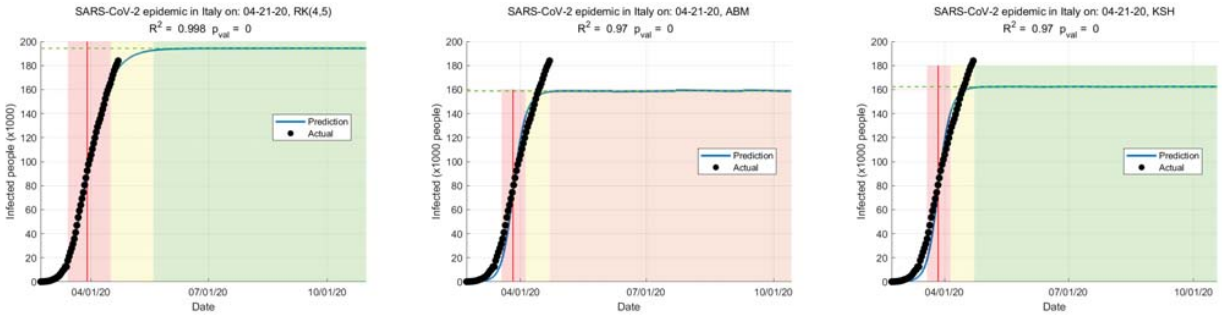

Fig. 19 Forecasting of SARS-CoV-2 epidemic by SIR model without vital dynamics for Italy

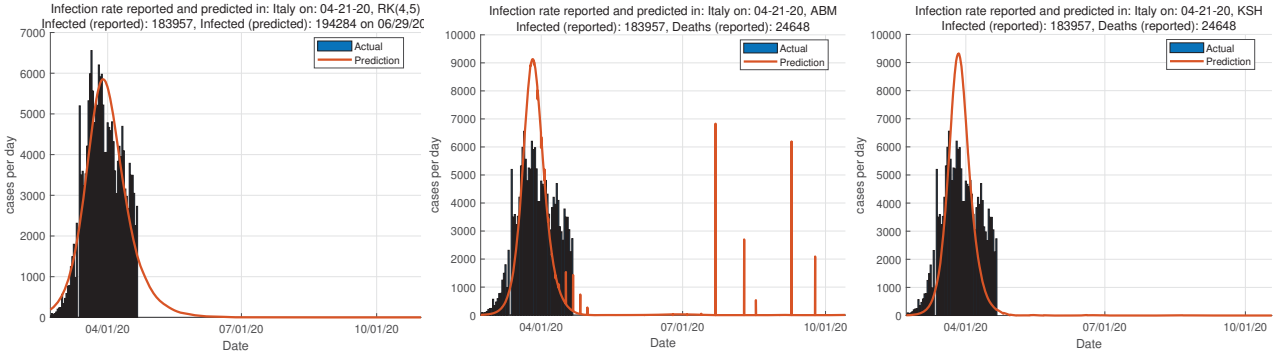

Fig. 20 The infection activity of SARS-CoV-2 predicted with the SIR model without vital dynamics for Italy

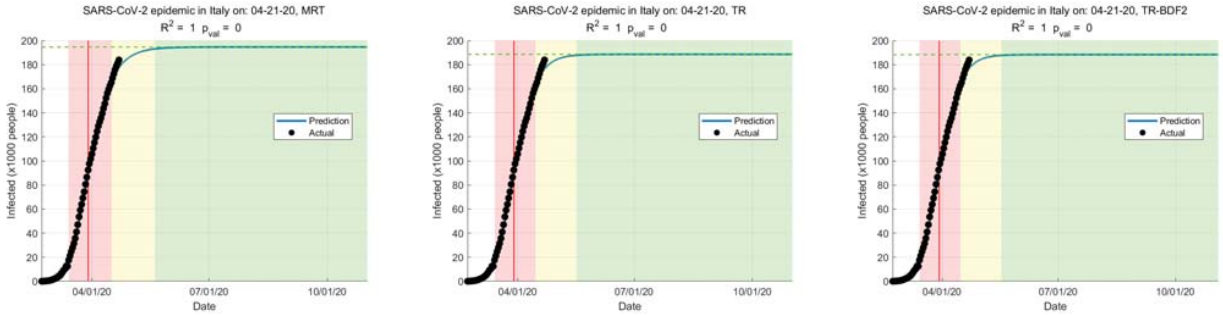

Fig. 21 Forecasting of SARS-CoV-2 epidemic by SIR model without vital dynamics for Italy

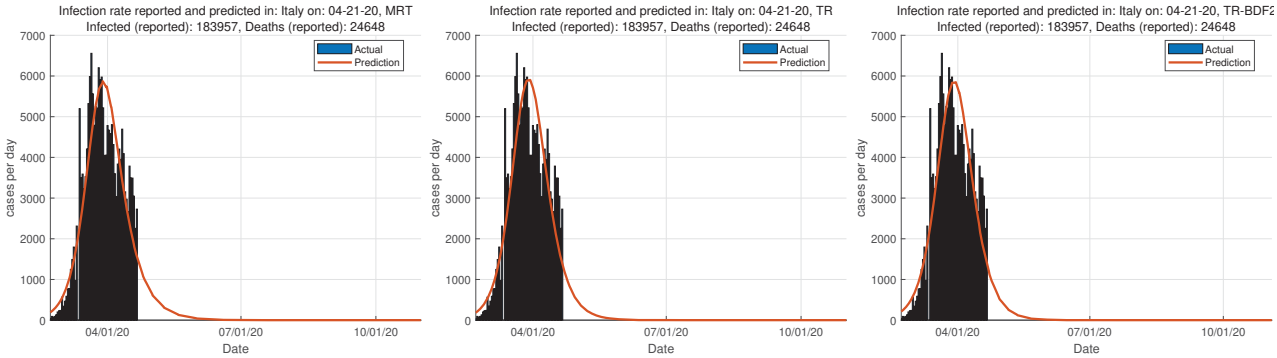

Fig. 22 The infection activity of SARS-CoV-2 predicted with the SIR model without vital dynamics for Italy

March 2020 the outbreak started to accelerate arriving at an inflection point on 06 April 2020, and on 24 April 2020 a stable growth starts and finally the ending phase comes on 31 May 2020. We see that from applying different numerical solutions, the ending phase finishes with almost one month difference.

From Fig. 31, it is predicted that on 31 May 2020, Denmark will have 681 deaths and 11331 recovered individuals.

Fig. 32 shows the transition rate from the compartment of susceptible individuals to the compartment of infectious individuals which passed the peak for Denmark and are now decreasing (for the new infected cases per day) and is on the way to a stable phase for the accumulated infected cases day after day.

While Fig. 33 indicates that the susceptible cases are less than 40000 cases and the unsusceptible cases more than 80000 cases. Fig. 34 shows a summary of the infected, dead, and recovered cases in order to better understand the active cases which are shown in red and correspond to the number of infected cases removing dead and recovered cases. On 21 April 2020, the predicted active cases for Denmark are at low rates.

Table III shows a summary of the most important epidemic indicators for all numerical solutions considered in the study for Denmark. The technique with smaller RMSE is the ABM method. The basic reproduction ratio varies from 1.07 to 1.53, which means that the secondary infections from one infected person are around one to two people. While the typical time between contacts varies from 0.59 to 3.08 days and the average infectious period varies from 0.63 to 3.35 days.

According to the RMSE for Denmark, the numerical

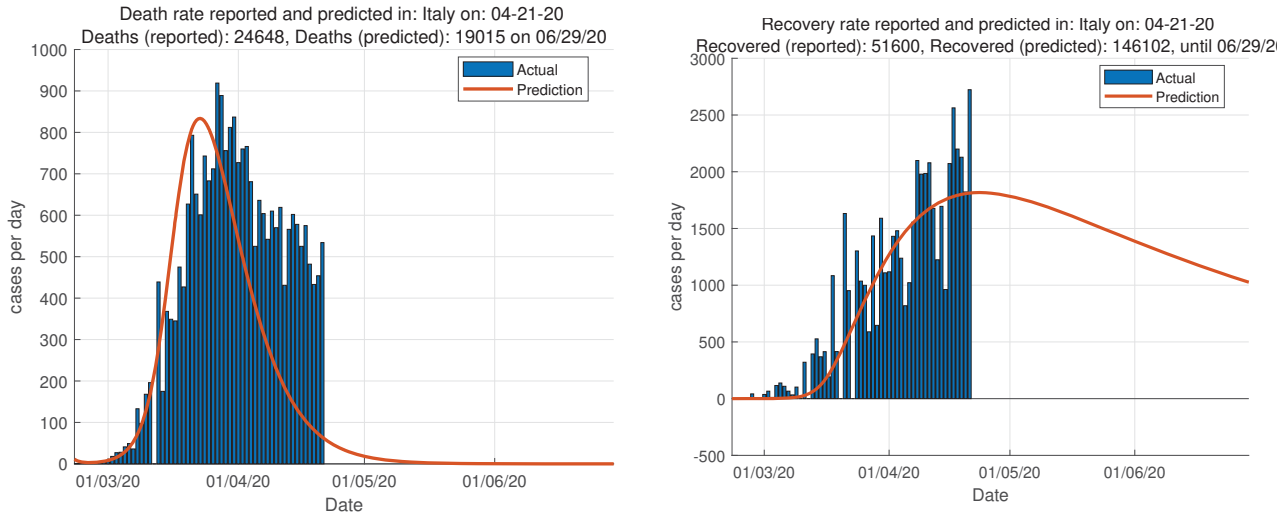

Fig. 23 Death and Recovery rates reported and predicted for Italy

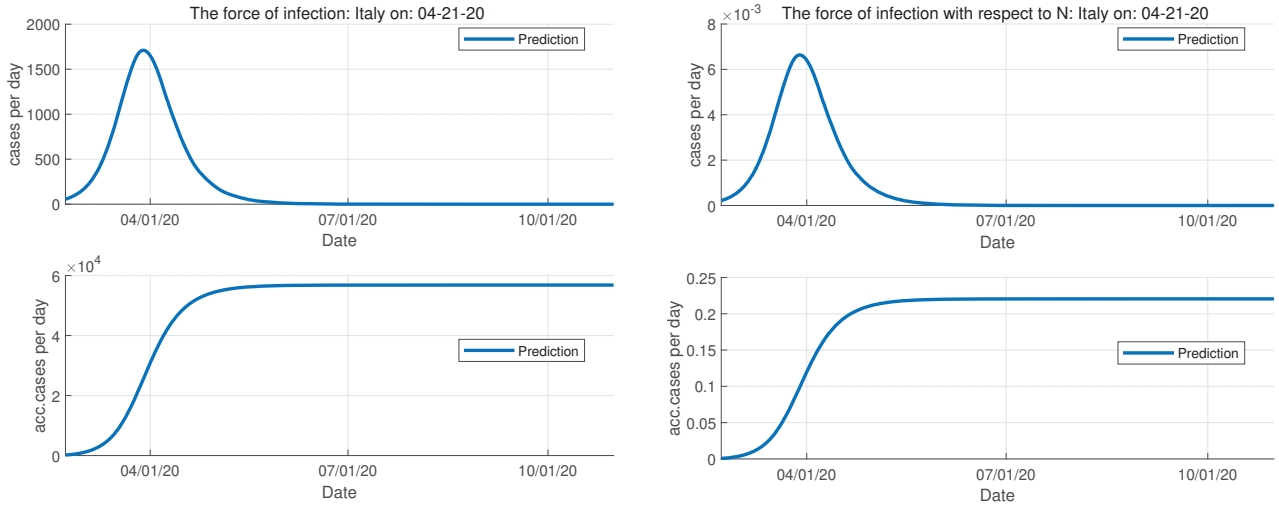

Fig. 24 The force of Infection which depends on the absolute number of infectious cases and on their fraction with respect to the total constant population for Italy

TABLE III
EPIDEMIC SIZE INDICATORS FOR DENMARK

Method	N	R_0	CP	IP	RMSE
RK(4,5)	62509	1.076	0.59	0.63	276.917
ABM	14115	1.7	1.96	3.35	161.281
KSH	14623	1.403	1.31	1.83	483.799
MRT	16223	1.529	3.08	4.7	170.167
TR	71735	1.071	0.66	0.71	174.768
TR-BDF2	22766	1.301	2.16	2.81	171.82

methods are listed (from the smallest to the highest value) as follows:

- ABM
- MRT
- TR-BDF2
- TR
- RK(4,5)
- KSH

Regarding Finland, from the chart produced by the ABM method (with a closer inspection), one can see that the outbreak started on 8 March 2020 and on 23 March 2020, the outbreak started to accelerate arriving at an inflection point on 6 April 2020, and on 22 April 2020 a stable growth starts and finally the ending phase comes on 21 May 2020.

Referring to RMSE, the ABM method is second listed with the smallest values, which is why we are going to interpret results of the MRT method identified as the best method with the smallest RMSE. Referring to the MRT method, one can see (with a closer inspection) that the outbreak in Finland started on 8 March 2020 and on 22 March 2020 the outbreak started to accelerate arriving in an inflection point on 10 April 2020, on 1 May 2020 a stable growth starts and finally the ending phase comes on 15 June 2020. We see that from applying different numerical solutions the ending phase finishes with almost one month difference.

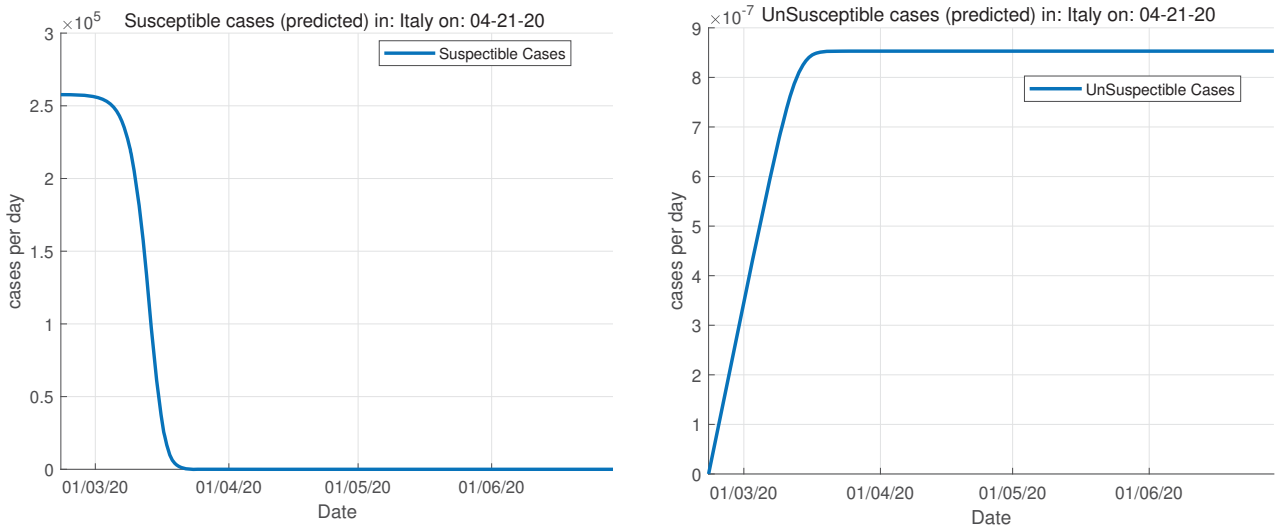

Fig. 25 All output channels for the SIR model without vital dynamics for Italy

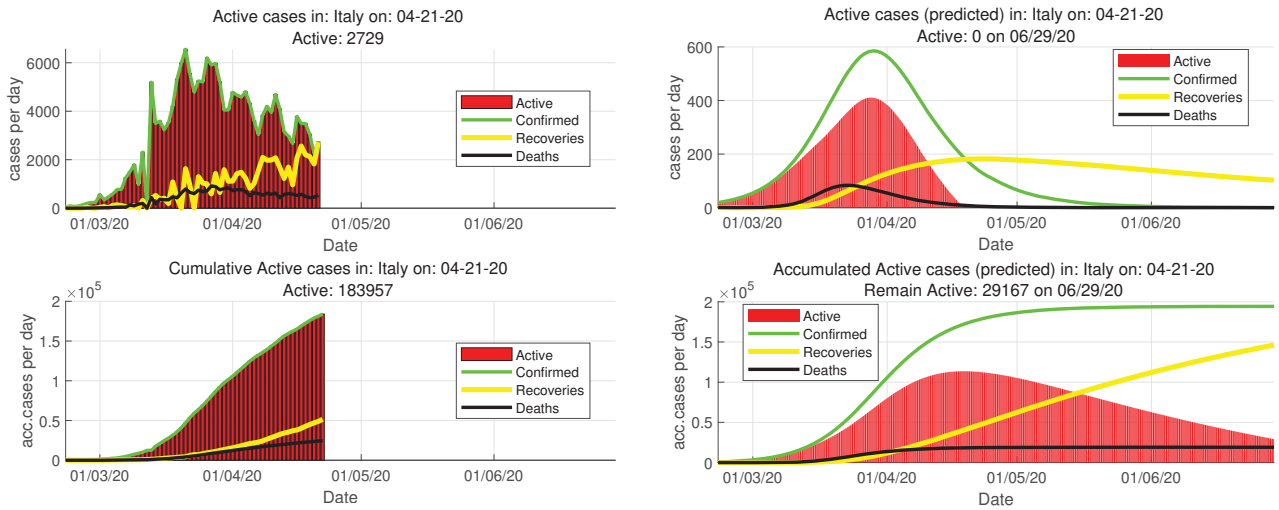

Fig. 26 Active cases reported and predicted from the SIR model without vital dynamics until the end of the first cycle of the activity of coronavirus for Italy

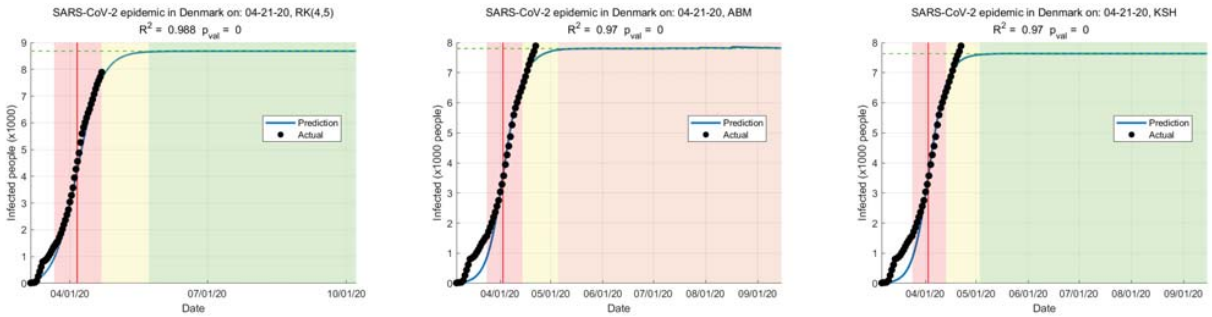

Fig. 27 Forecasting of SARS-CoV-2 epidemic by the SIR model without vital dynamics for Denmark

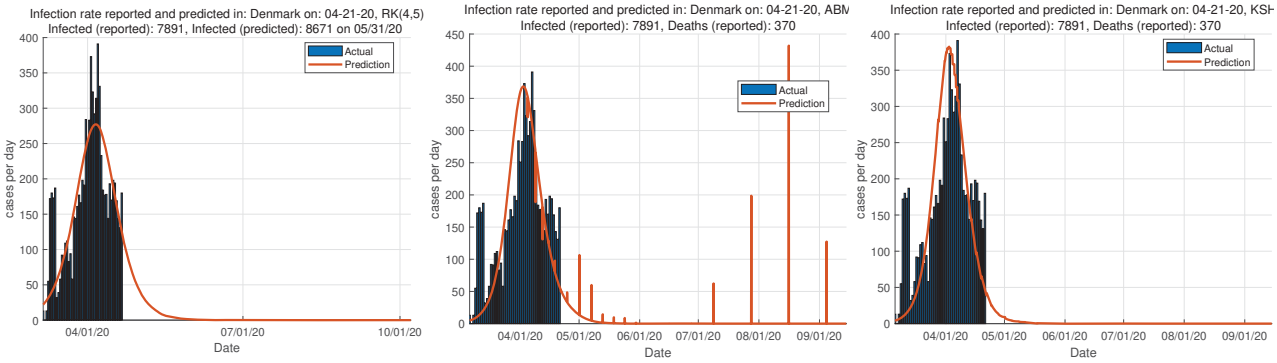

Fig. 28 The infection activity of SARS-CoV-2 predicted with the SIR model without vital dynamics for Denmark

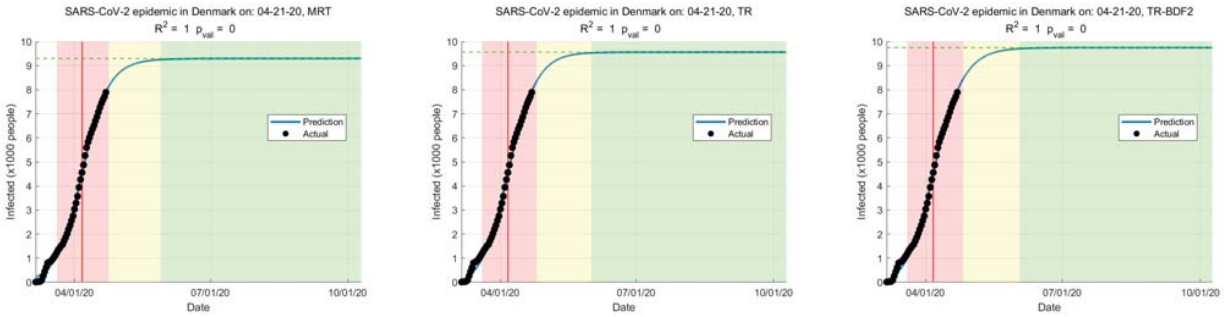

Fig. 29 Forecasting of SARS-CoV-2 epidemic by the SIR model without vital dynamics for Denmark

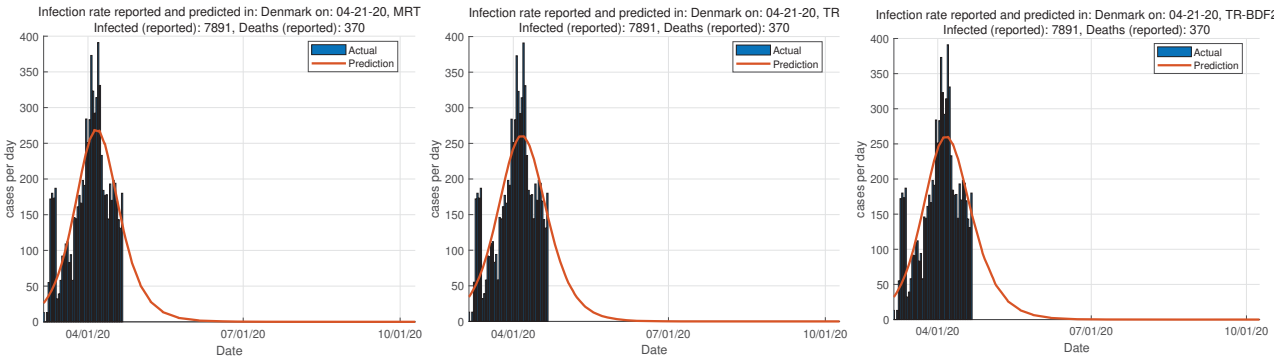

Fig. 30 The infection activity of SARS-CoV-2 predicted with SIR model without vital dynamics for Denmark

From Fig. 39 on 15 June 2020, Finland will have 256 deaths and 7017 recovered individuals.

Fig. 40 displays the transition rate from the compartment of susceptible individuals to the compartment of infectious individuals which passed the peak for Finland and now is decreasing for the new infected cases per day and is on its way to the stable growth phase for the accumulated infected cases day after day. While Fig. 41 indicates that the susceptible cases at the end of the ending phase will be less then 12000 cases and the unsusceptible cases more than 3000 cases. Fig. 42 shows a summary of the infected, dead, and recovered cases in order to better understand the active cases which are shown in red and corresponds to the number of infected cases removing dead and recovered cases. On 21 April 2020, the predicted active cases for Finland are at low rates.

Table IV shows the summary of the epidemic indicators for all numerical solution considered in the study for Finland. The technique with smaller RMSE is the ABM method. The basic reproduction ratio varies from 1.095 to 1.369, which means that the secondary infection from one infected person is around one person. While the typical time between contacts varies from 0.9 to 2.62 days and the average infectious period varies from 0.87 to 4.34 days. According to the RMSE for Finland, the numerical methods are listed (from the smallest to the highest value) as follows:

- MRT
- ABM
- RK(4,5)
- TR-BDF2
- TR

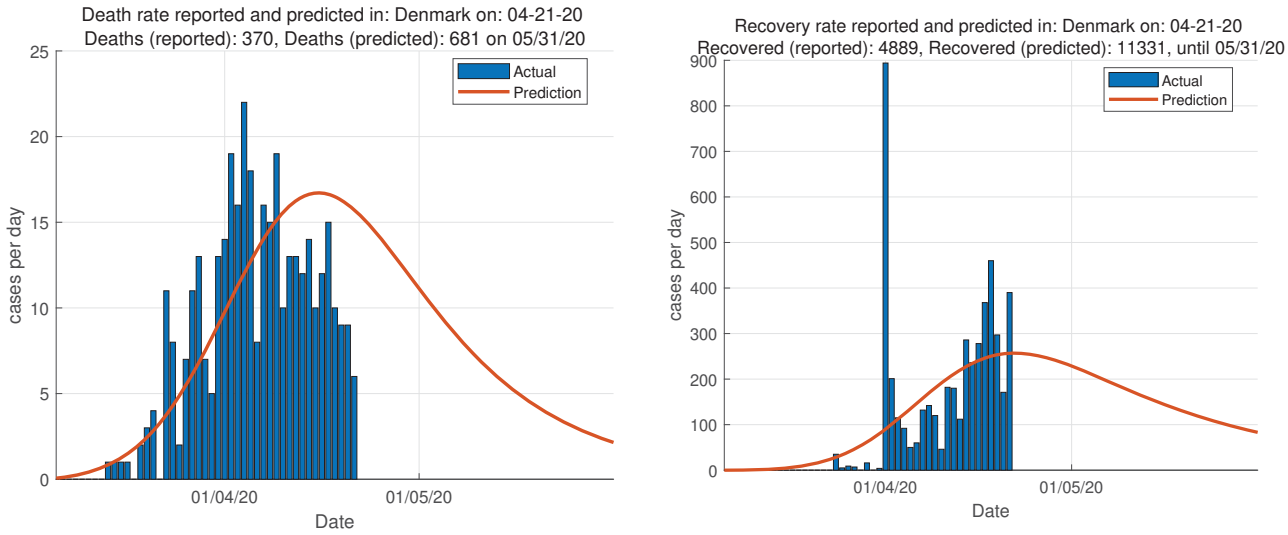

Fig. 31 Death and Recovery rates reported and predicted for Denmark

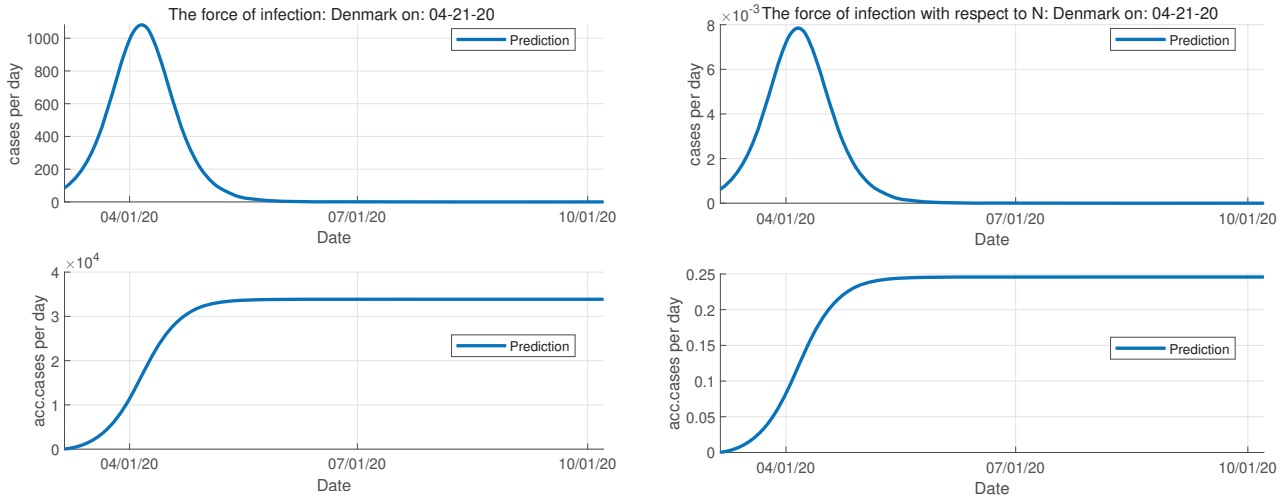

Fig. 32 The force of Infection which depend on the absolute number of infectious cases and on their fraction with respect to the total constant population for Denmark

- KSH

TABLE IV
EPIDEMIC INDICATORS FOR FINLAND

Method	N	R_0	CP	IP	RMSE
RK(4,5)	29482	1.095	0.9	0.99	56.4876
ABM	12635	1.278	2.1	2.69	56.4708
KSH	6369	1.661	2.62	4.34	116.522
MRT	33225	1.083	0.8	0.87	56.3914
TR	10512	1.369	2.54	3.47	59.01
TR-BDF2	22467	1.133	1.2	1.36	56.972

In the case of Norway, from the chart produced by ABM method (from a closer inspection), one can see that the outbreak in Norway started on 4 March 2020 and on 14 March 2020, the outbreak started to accelerate arriving at an inflection point on 27 March 2020, on 12 April 2020 a stable growth starts and finally the ending phase comes on 21 May 2020. Referring to RMSE, the ABM method is the best method for the smallest RMSE but we are going to interpret the results of the MRT method, identified as second listed with the smallest RMSE. Referring to the MRT method, one can see (with closer inspection) that the outbreak in Norway started on 4 March 2020 and on 14 March 2020, the outbreak started to accelerate arriving at an inflection point on 27 March 2020, and on 12 April 2020 a stable growth starts and finally the ending phase comes on 11 May 2020. We see that from applying different numerical solutions the ending phase finishes almost in the same time.

From Fig. 47, on 21 May 2020, Norway will have 215 deaths and 328 recovered individuals. Fig. 48 displays the transition rate from the compartment of susceptible individuals to the compartment of infectious individuals which passed the peak for Norway and is less than 100 new infected cases per day and is on the way to the stable growth phase for the accumulated infected cases day after day. While Fig. 49

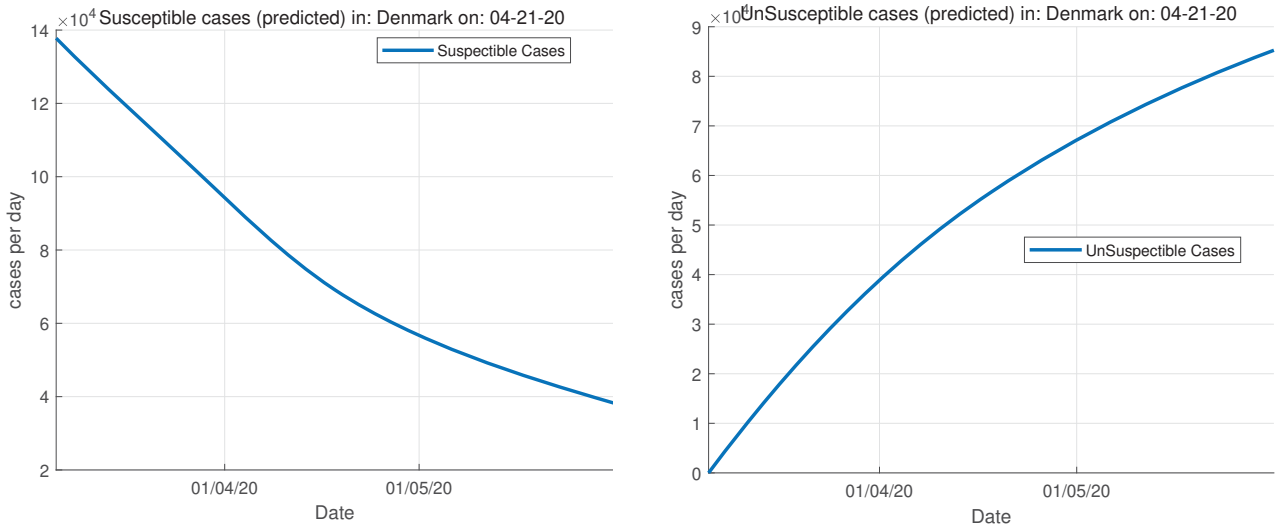

Fig. 33 All output channels for the SIR model without vital dynamics for Denmark

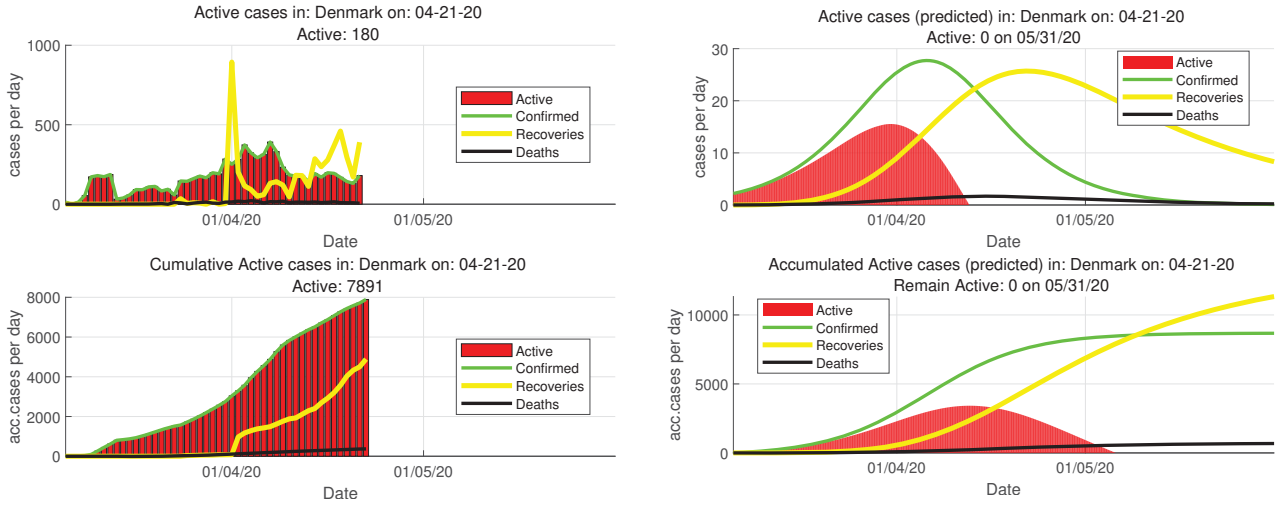

Fig. 34 Active cases reported and predicted from the SIR model without vital dynamics until the end of the first cycle of the activity of coronavirus for Denmark

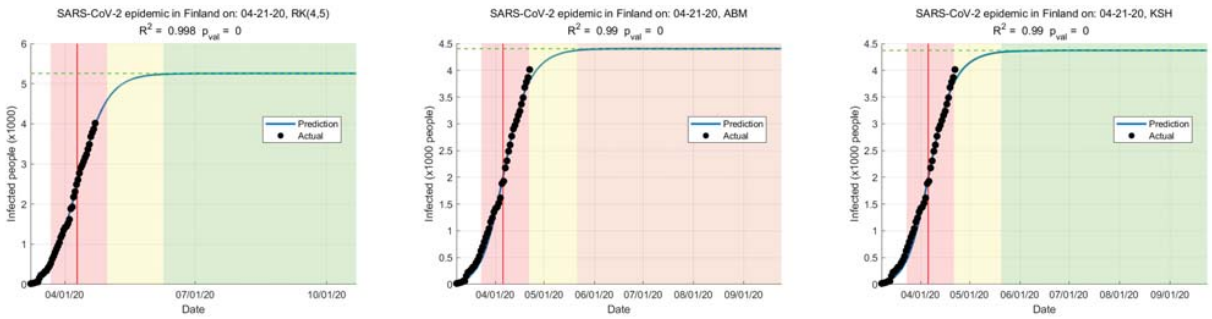

Fig. 35 Forecasting of SARS-CoV-2 epidemic by SIR model without vital dynamics for Finland

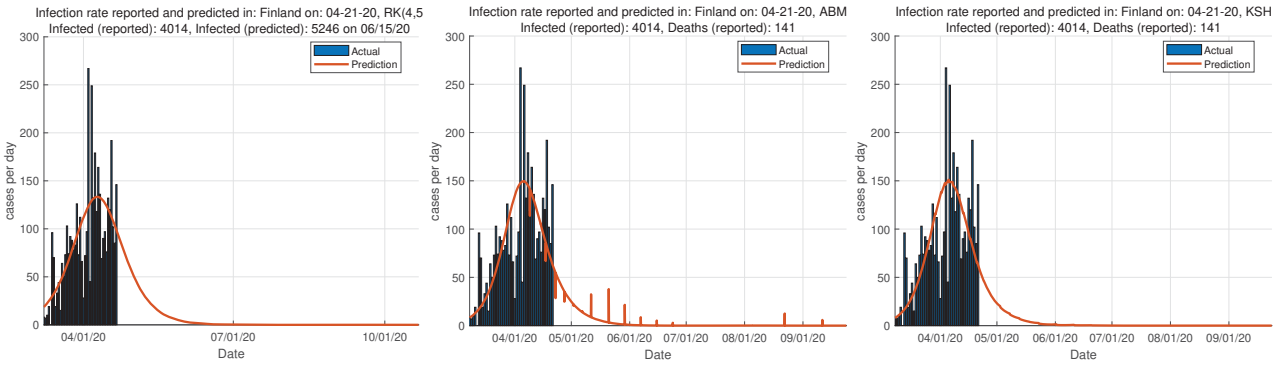

Fig. 36 The infection activity of SARS-CoV-2 predicted with the SIR model without vital dynamics for Finland

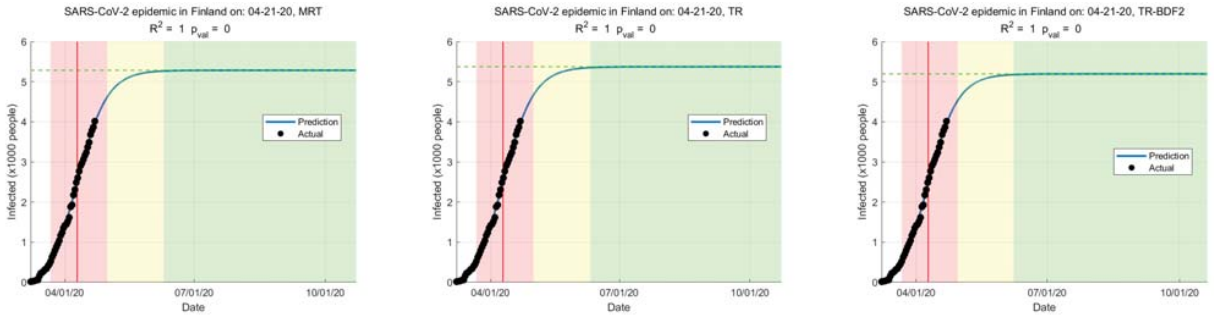

Fig. 37 Forecasting of SARS-CoV-2 epidemic by the SIR model without vital dynamics for Finland

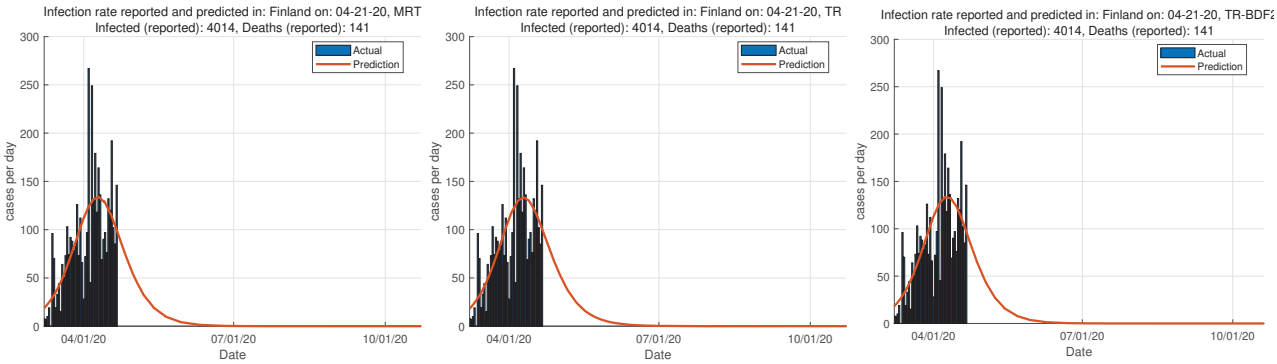

Fig. 38 The infection activity of SARS-CoV-2 predicted with the SIR model without vital dynamics for Finland

indicates that the susceptible cases at the end of the ending phase will be almost zero and the unsusceptible cases will be more than 22000 cases. Fig. 50 shows a summary of the infected, dead, and recovered cases in order to better understand the active cases which are shown in red and corresponds to the number of infected cases removing dead and recovered cases. On 21 April 2020, the predicted active cases for Norway are at low rates.

Table V shows a summary of the epidemic indicators for all numerical solution considered in the study for Norway. The technique with smaller RMSE is the ABM method. The basic reproduction ratio varies from 1.209 to 2.012, which means that the secondary infections from one infected person are around one to two people. While the typical time between contacts varies from 0.85 to 2.76 days and the average infectious period varies from 0.95 to 5.55 days.

TABLE V
EPIDEMIC SIZE INDICATORS FOR NORWAY

Method	N	R_0	CP	IP	RMSE
RK(4,5)	22182	1.209	1.29	1.56	84.2228
ABM	12862	1.485	2.28	3.39	82.6971
KSH	8935	2.012	2.76	5.55	152.565
MRT	33725	1.123	0.85	0.95	83.2052
TR	11602	1.596	2.59	4.14	88.3927
TR-BDF2	13110	1.466	2.21	3.24	86.6238

According to the RMSE for Norway, the numerical methods are sorted (from the smallest to the highest value) as follows:

- ABM
- MRT
- RK(4,5)

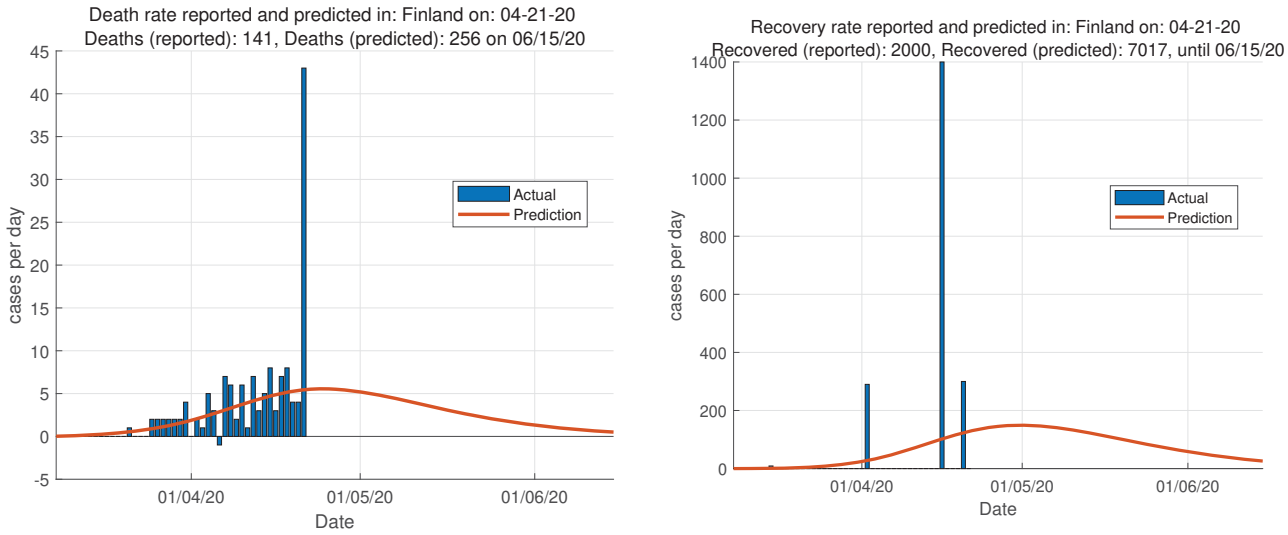

Fig. 39 Death and Recovery rates reported and predicted for Finland

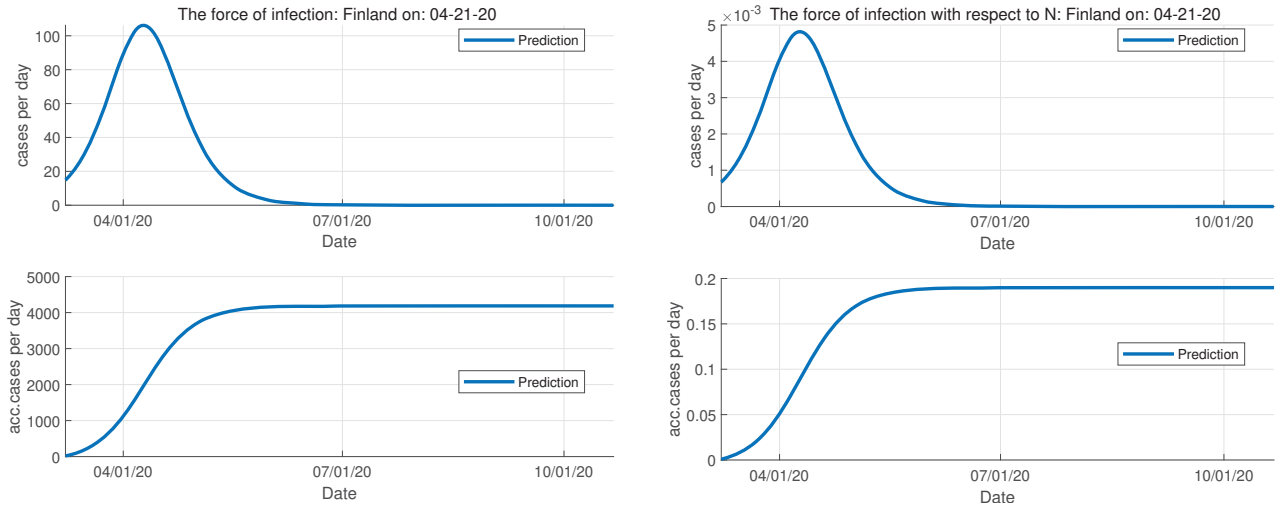

Fig. 40 The force of Infection which depend on the absolute number of infectious cases and on their fraction with respect to the total constant population for Finland

- TR-BDF2
- TR
- KSH

Regarding Sweden, from the chart produced by the ABM method (with closer inspection), one can see that the outbreak started on 4 March 2020 and on 26 March 2020 the outbreak started to accelerate arriving at an inflection point on 10 April 2020, and on 28 April 2020 a stable growth starts and finally the ending phase comes on 31 May 2020. Referring to RMSE, the ABM method is the best method for the smallest RMSE but we are going to interpret results of the MRT method identified as second listed with the smallest RMSE. Referring to the MRT method, one can see (from closer inspection) that the outbreak in Sweden started on 4 March 2020 and on 26 March 2020 the outbreak started to accelerate arriving at an inflection point on 16 April 2020, and on 10 May 2020 a stable growth starts and finally the ending phase comes on 23 June 2020. We see that from applying different numerical solutions, the ending phase finishes with almost one month difference.

From Fig. 55, it is predicted that on 2 July 2020, Sweden will have 6113 deaths and 3036 recovered individuals. Fig. 56 displays the transition rate from the compartment of susceptible individuals to the compartment of infectious individuals which passed the peak for Sweden and still is in high rates but in a decreasing mode (more than 200) for the new infected cases per day. While Fig. 57 indicates that the susceptible cases will be less than 20000 at the end of the ending phase and the unsusceptible cases more than 18000 cases. Fig. 58 shows a summary of the infected, dead, and recovered cases in order to better understand the active cases which are shown in red and corresponds to the number of infected cases removing dead and recovered cases. On 21 April

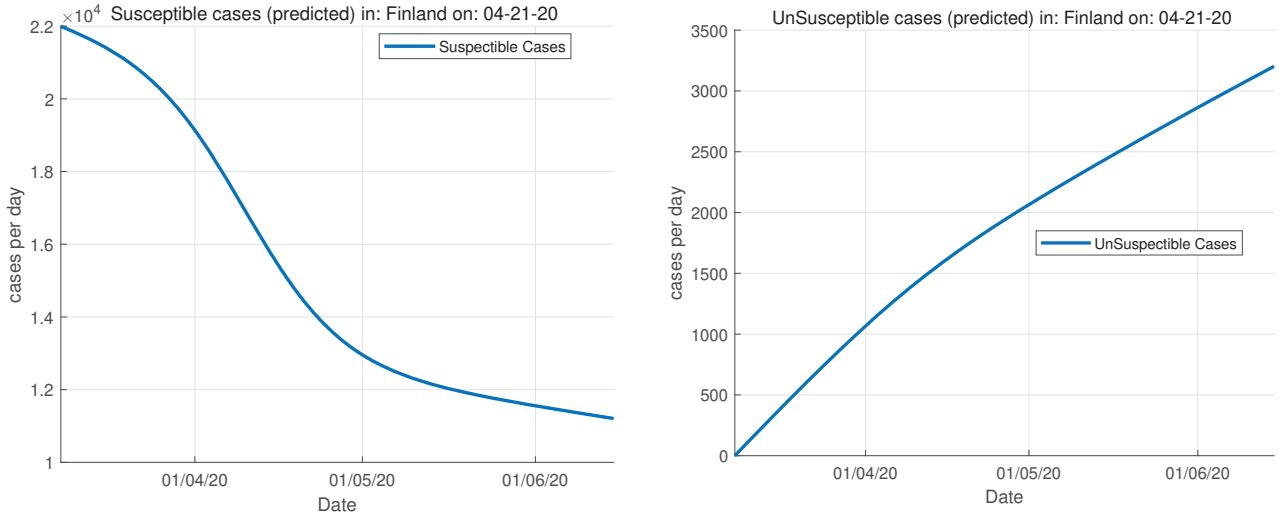

Fig. 41 All output channels for the SIR model without vital dynamics for Finland

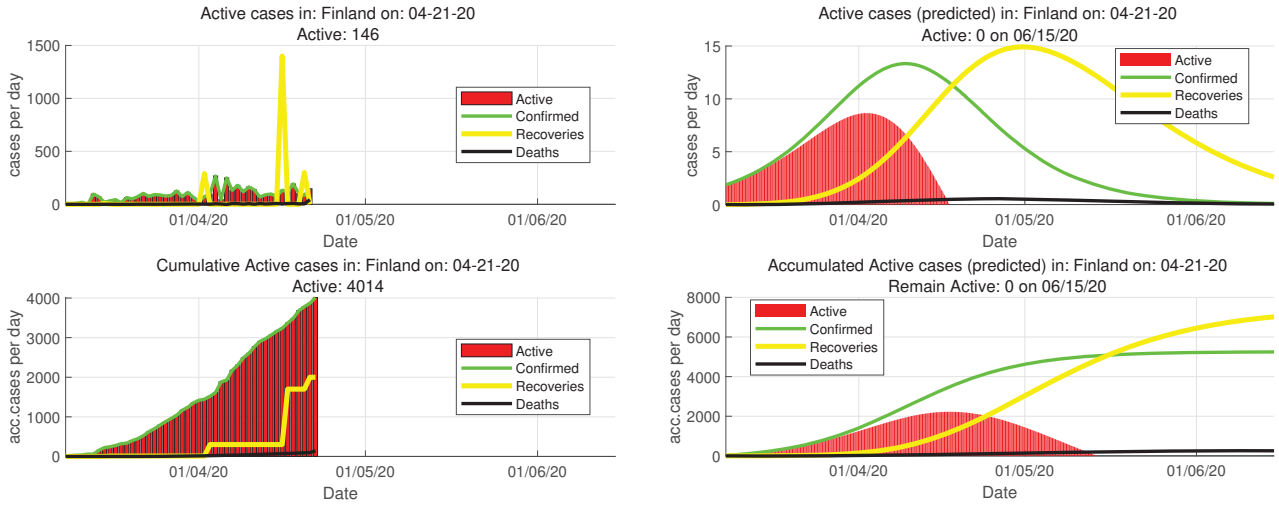

Fig. 42 Active cases reported and predicted from the SIR model without vital dynamics until the end of the first cycle of the activity of coronavirus for Finland

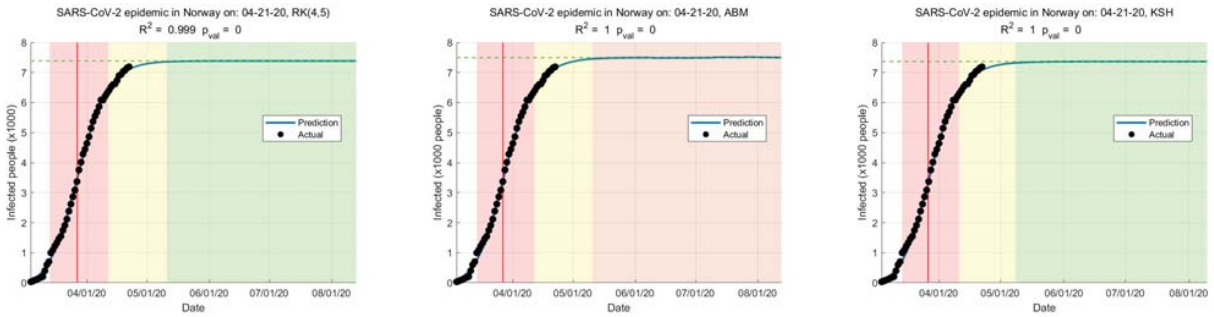

Fig. 43 Forecasting of SARS-CoV-2 epidemic by the SIR model without vital dynamics for Norway

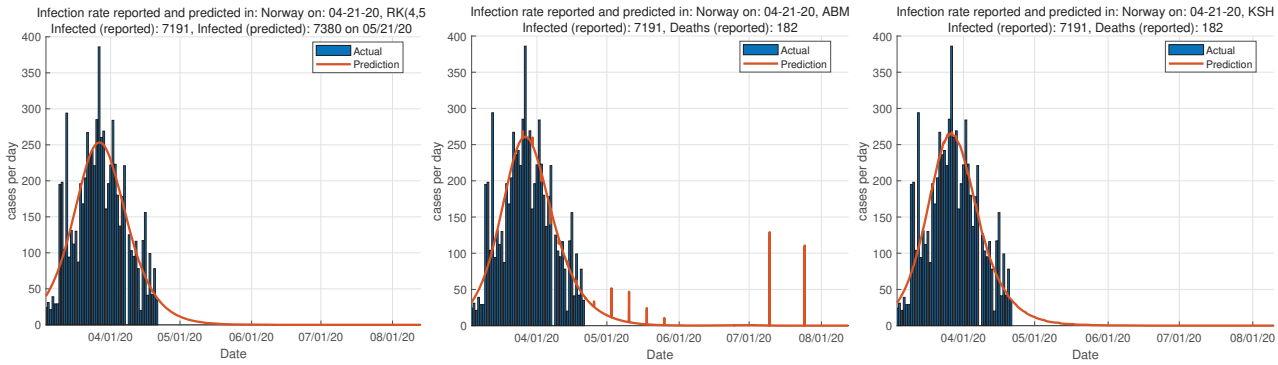

Fig. 44 The infection activity of SARS-CoV-2 predicted with the SIR model without vital dynamics for Norway

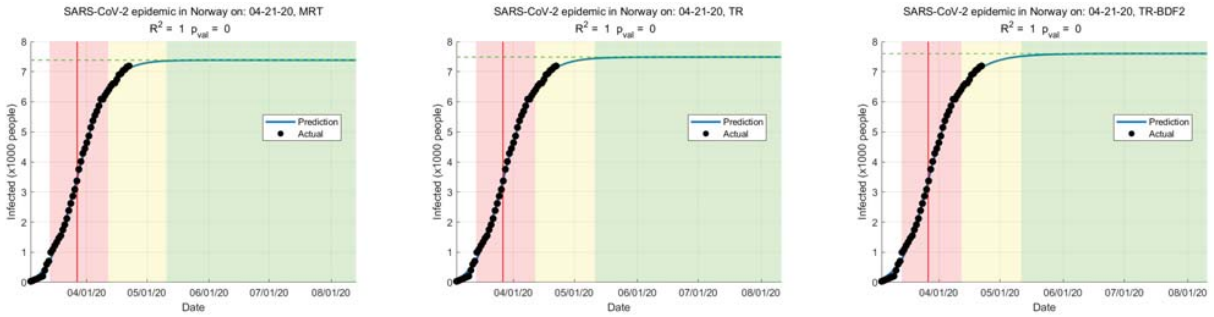

Fig. 45 Forecasting of SARS-CoV-2 epidemic by the SIR model without vital dynamics for Norway

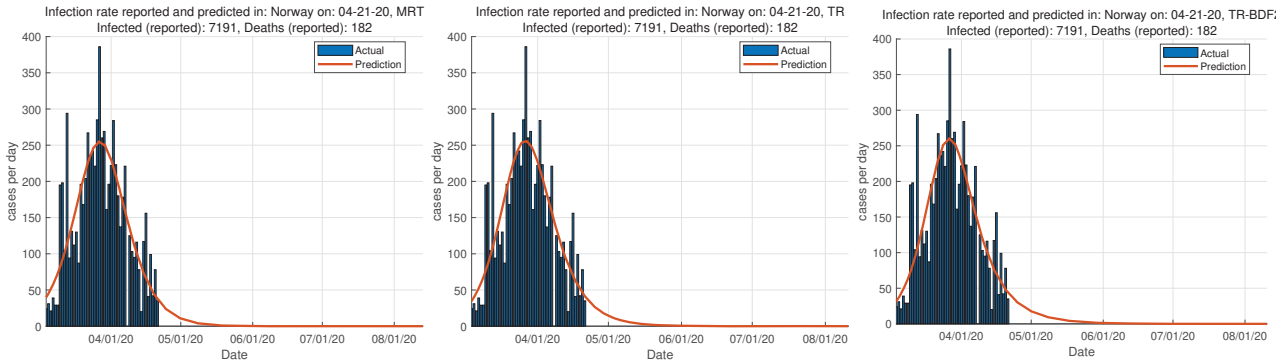

Fig. 46 The infection activity of SARS-CoV-2 predicted with the SIR model without vital dynamics for Norway

2020, the predicted active cases for Sweden are lower than before but still at high rates.

Table VI shows a summary of the epidemic indicators for all numerical solution considered in the study for Sweden. From the table, we see that the technique with smaller RMSE is the ABM method. The basic reproduction ratio varies from 1.28 to 1.92, which means that the secondary infections from one infected person are around one to two people. While the typical time between contacts is varying from 1.53 to 3.36 days and the average infectious period varies from 2.09 to 6.02 days.

According to the RMSE for Sweden, the numerical methods are listed (from the smallest to the highest value) as follows:

- ABM
- MRT
- RK(4,5)
- TR-BDF2
- KSH
- TR

Note that all the results change over time with the new data added to the dataset. From the outputs it is clear that

TABLE VI
EPIDEMIC SIZE INDICATORS FOR SWEDEN

Method	N	R_0	CP	IP	RMSE
RK(4,5)	48000	1.338	2.41	3.22	134.03
ABM	22212	1.922	3.13	6.02	103.459
KSH	31394	1.383	1.56	2.15	581.486
MRT	53497	1.286	2.14	2.75	134.022
TR	32248	1.369	1.53	2.09	586.748
TR-BDF2	35436	1.572	3.36	5.29	134.925

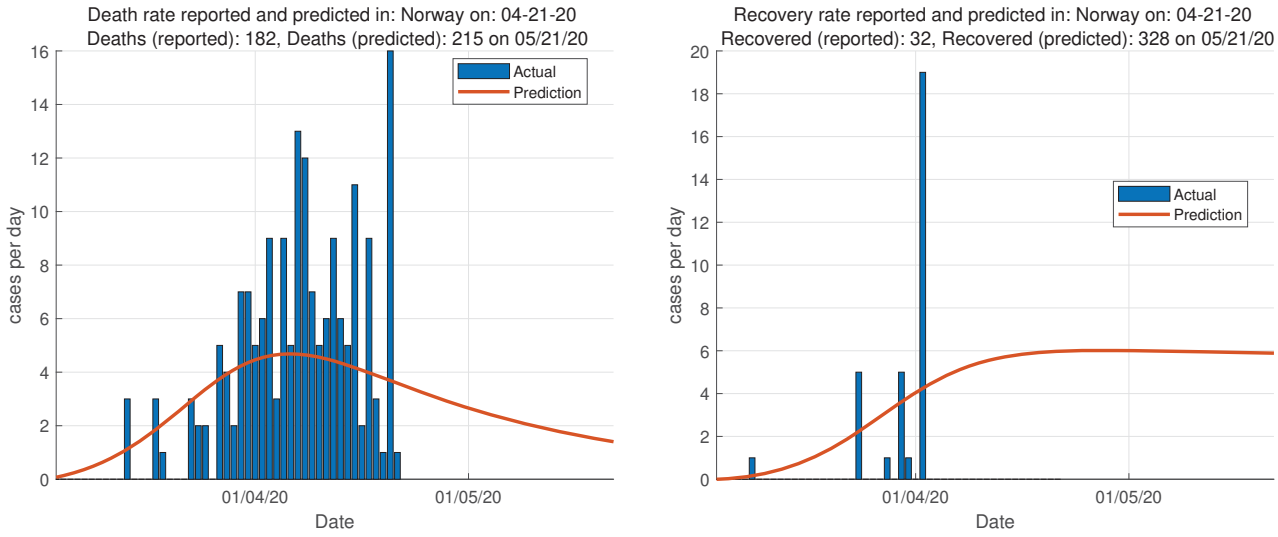

Fig. 47 Death and Recovery rates reported and predicted for Norway

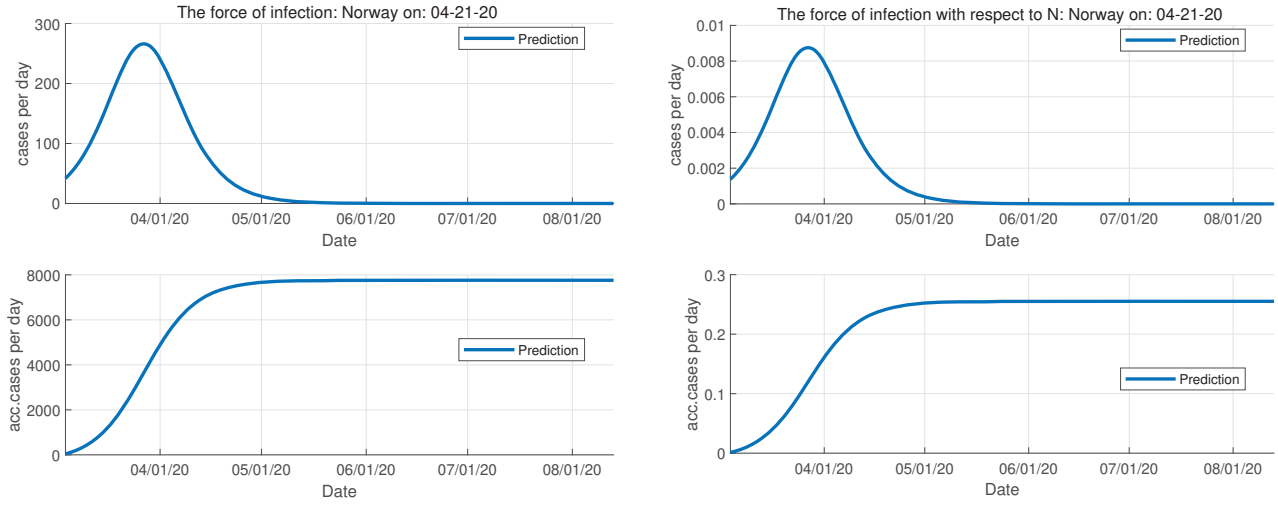

Fig. 48 The force of Infection which depend on the absolute number of infectious cases and on their fraction with respect to the total constant population for Norway

predictor-corrector methods (ABM method) perform better than other methods; however, from our analysis the MRT method is another method that fits well in our data. Comparing with the ABM method, the MRT method is weak for tracing the activity of the virus after the ending phase of the first course of the virus. Improvements need to be done in order to obtain more accurate results and always considering the fact that all models are wrong but some are useful (Box (1976)). In the appendix is found some more information about the continuation of the virus activity after the ending phase in Nordic countries with the ABM method. It is possible to see when a reactivation of the virus at high rates will come.

VII. Conclusions

SARS-CoV-2 is responsible for the pandemic of a respiratory disease spreading from person-to-person which has been officially announced by WHO on March 11. Different countries (even different cities inside a country) are facing with different levels of SARS-CoV-2 activity. The duration and severity of each pandemic phase can vary depending on the characteristics of the virus and the public health response. Note that Norway changed its preventative measures on 19 April 2020, and the prediction of pandemic is directly related with the softening/strengthening of the prevention measures and public behavior. Therefore, some changes in the infection curve of Norway is expected. Referring to the analysis conducted in this study on phylogenetic tree, it is found that the source of the SARS-CoV-2 epidemic is the bat. China and Italy were the roots of the spread for most of the international epidemic and the virus arrived in Nordic countries through Italy and the USA. In addition, a comparison of numerical methods for solving ODEs is done, where

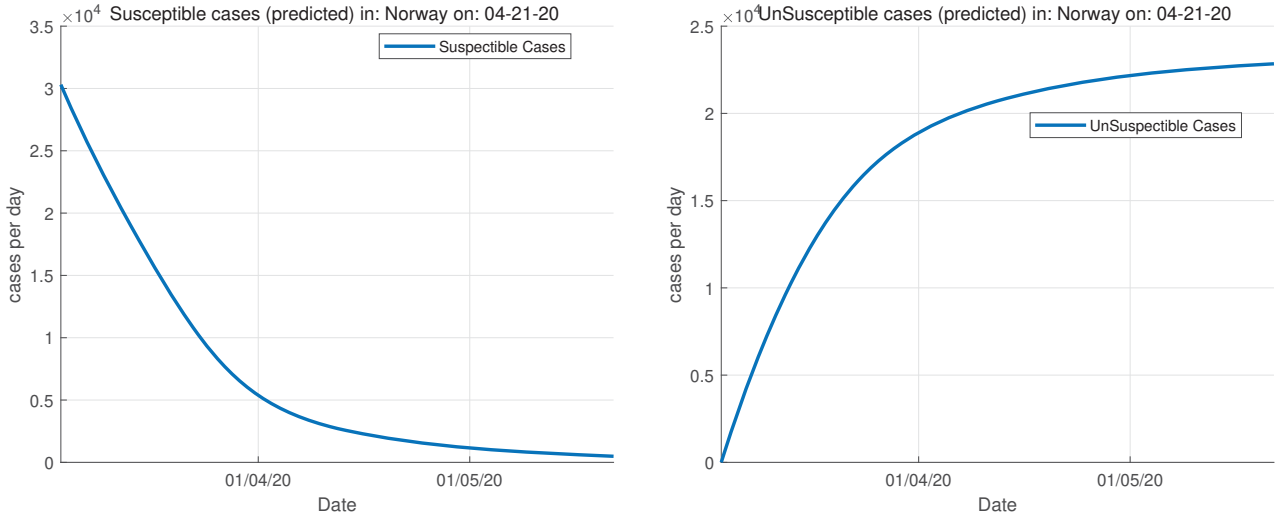

Fig. 49 All output channels for the SIR model without vital dynamics for Norway

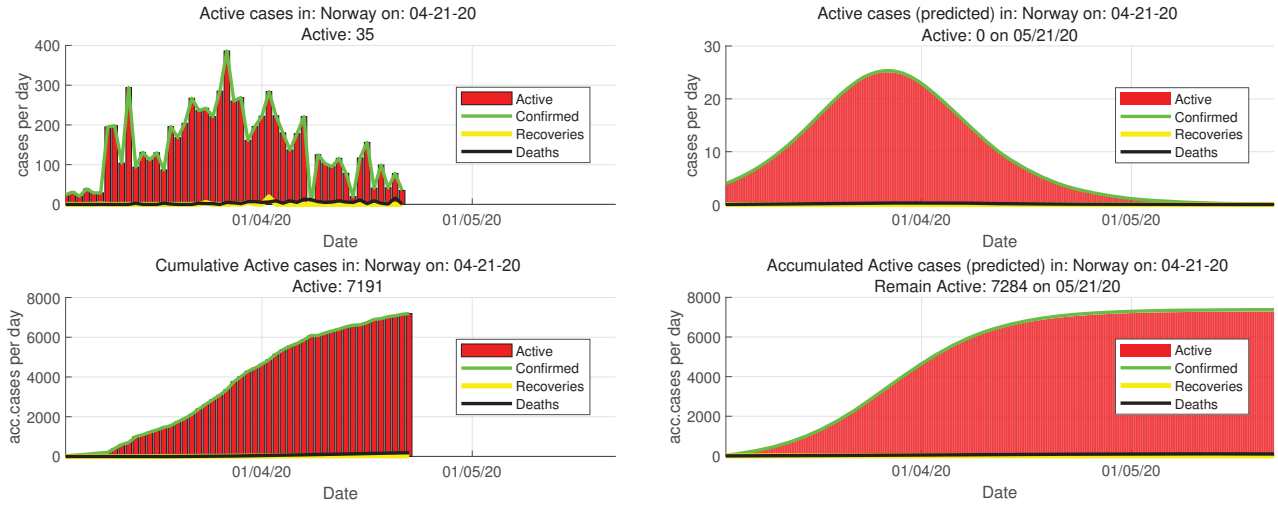

Fig. 50 Active cases reported and predicted from the SIR model without vital dynamics until the end of the first cycle of the activity of coronavirus for Norway

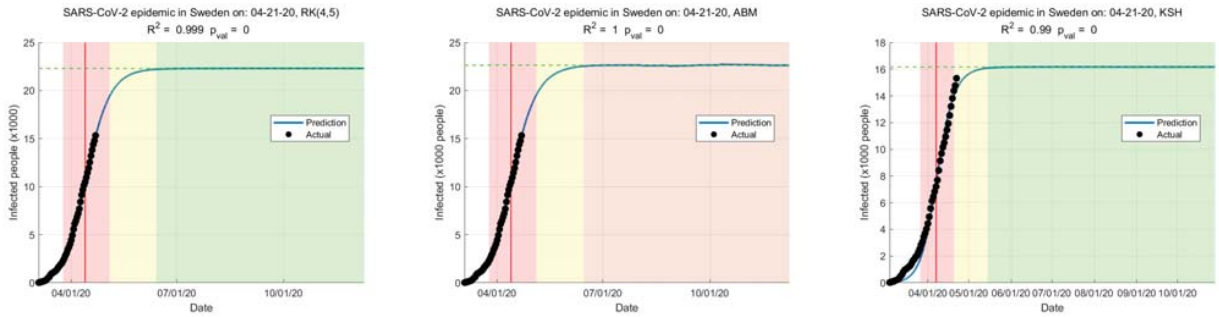

Fig. 51 Forecasting of SARS-CoV-2 epidemic by the SIR model without vital dynamics for Sweden

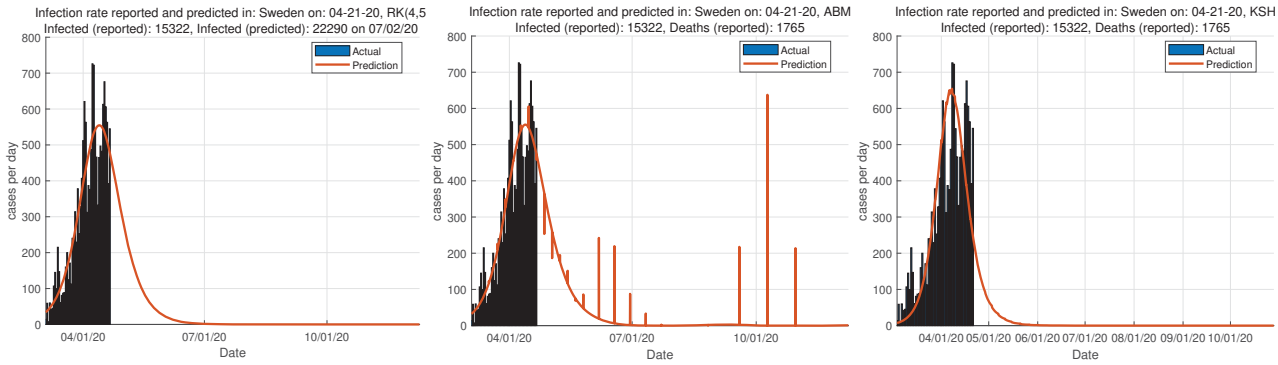

Fig. 52 The infection activity of SARS-CoV-2 predicted with the SIR model without vital dynamics for Sweden

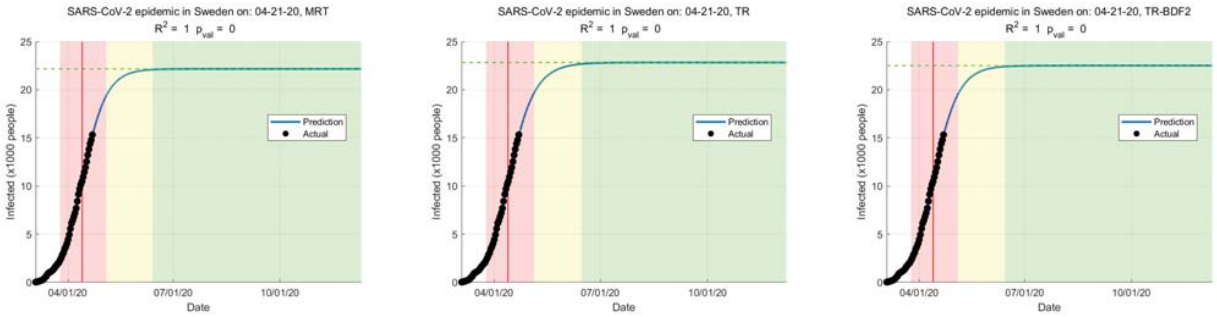

Fig. 53 Forecasting of SARS-CoV-2 epidemic by the SIR model without vital dynamics for Sweden

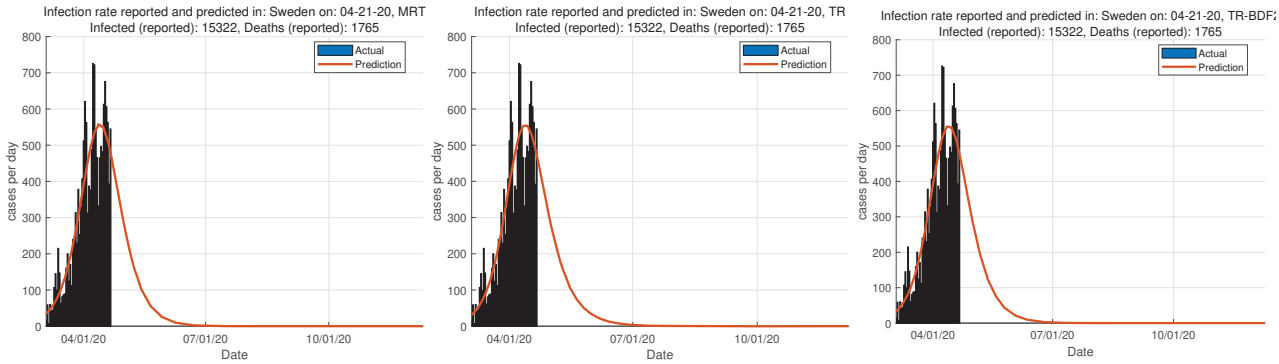

Fig. 54 The infection activity of SARS-CoV-2 predicted with the SIR model without vital dynamics for Sweden

predictor-corrector methods are estimated as more accurate. In this study, a SIR model is developed using six different numerical techniques in solving ODEs. By this model, we analyzed and predicted the diffusion of SARS-CoV-2 disease. We predicted the number of confirmed cases of SARS-CoV-2 for the period after 21 April 2020; the period of time when the virus will be in the ending phase of its course, and we looked for any possible return of virus activity during the summer and beyond. Note that the SIR model is a data- and method-driven model, therefore its accuracy will be increased over time. All numerical techniques display different dates to explain the different phases of the epidemic, except for Norway where the results from all methods seems to be in the same line. It is predicted that the course of the epidemic in Denmark will end on 2 June 2020, for Norway on 11 May 2020, for Finland on 10 June 2020 and for Sweden on 14 June 2020. Note that these results do not come from the method which was performing better for one country but from the method which gave the furthest ending date for that country. Then, the value of the basic reproduction number was computed to be $(1.07 - 1.53)$ for Denmark, $(1.1 - 1.37)$ for Finland, $(1.21 - 2.01)$ for Norway and $(1.28 - 1.92)$ for Sweden. Moreover, during performance evaluation, our model computed the value of RMSE for the ABM method to be the smallest one but MRT method was almost at the same level. The results obtained from this study are taken from training data up to 21 April 2020. Further, looking at the trend, there is definitely going to be a decrease in the number of cases for all Nordic countries. Actually, from the predictions, Norway and Denmark passed the acceleration phase and the inflection

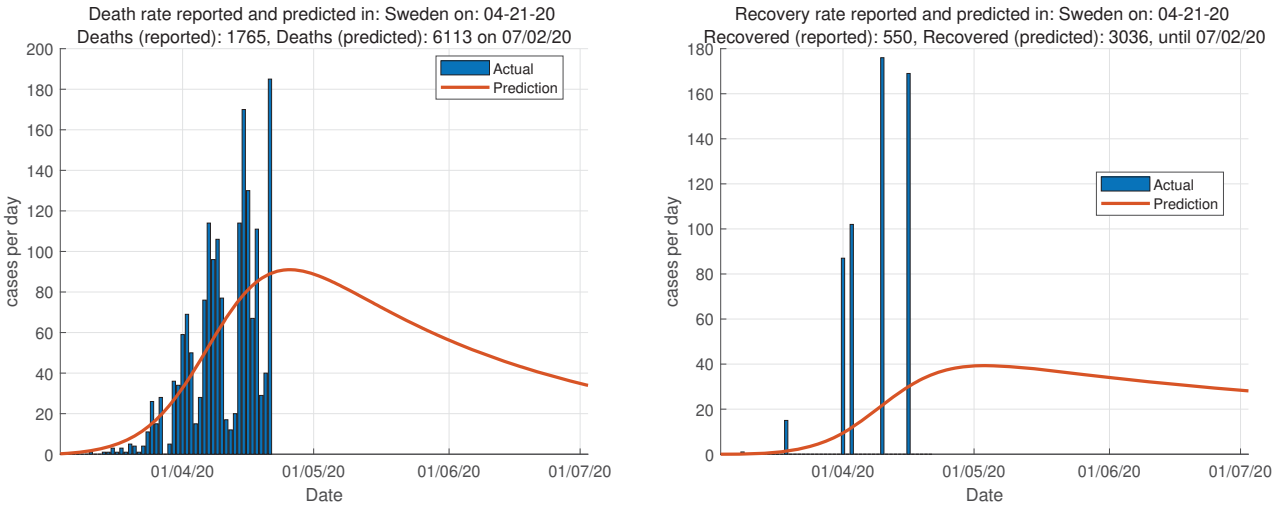

Fig. 55 Death and Recovery rates reported and predicted for Sweden

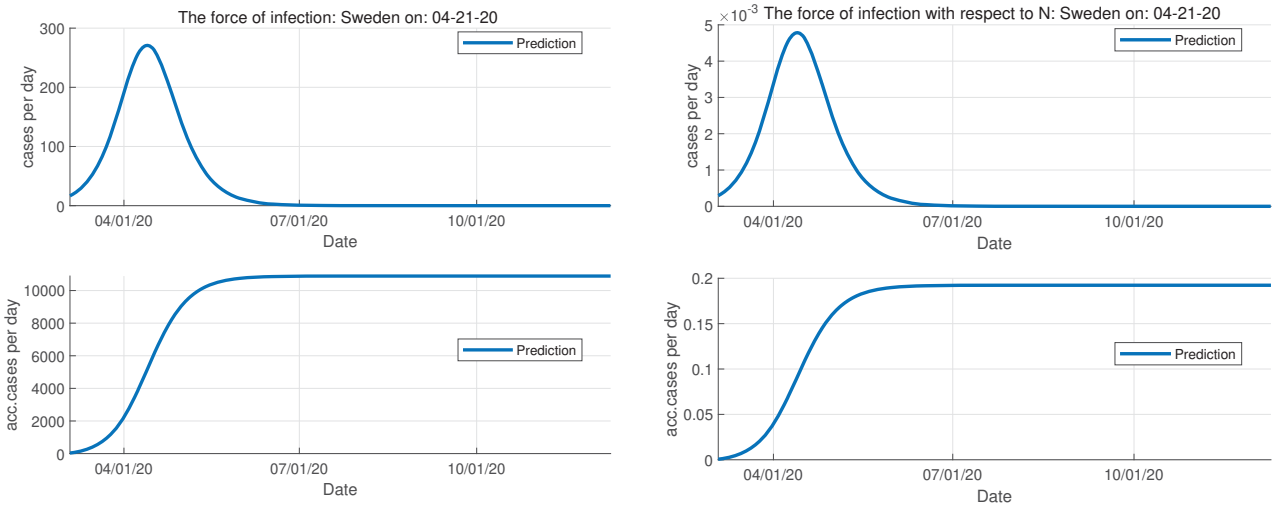

Fig. 56 The force of Infection which depend on the absolute number of infectious cases and on their fraction with respect to the total constant population for Sweden

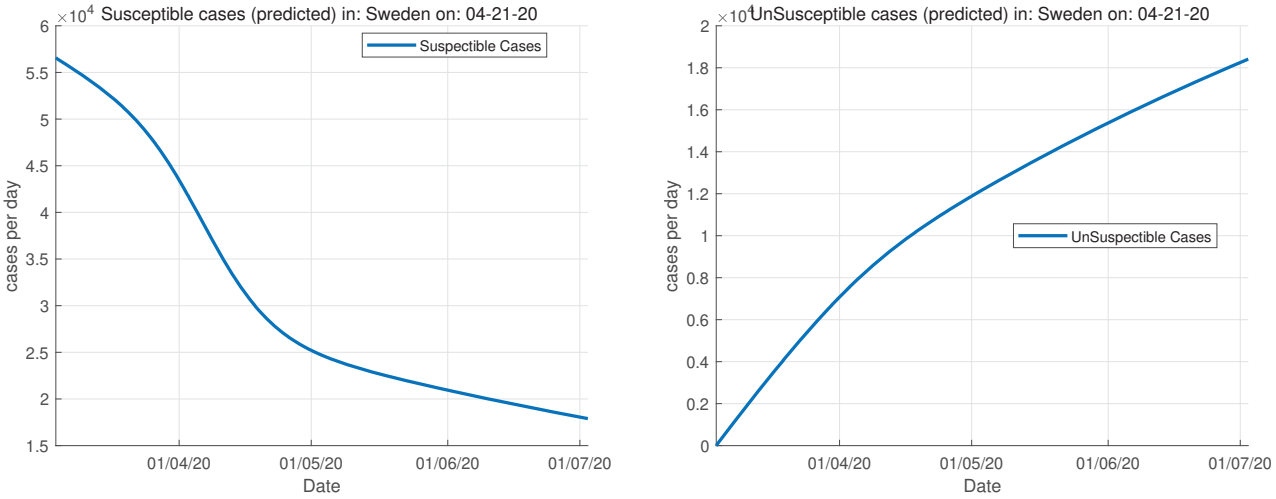

Fig. 57 All output channels for SIR model without vital dynamics for Sweden

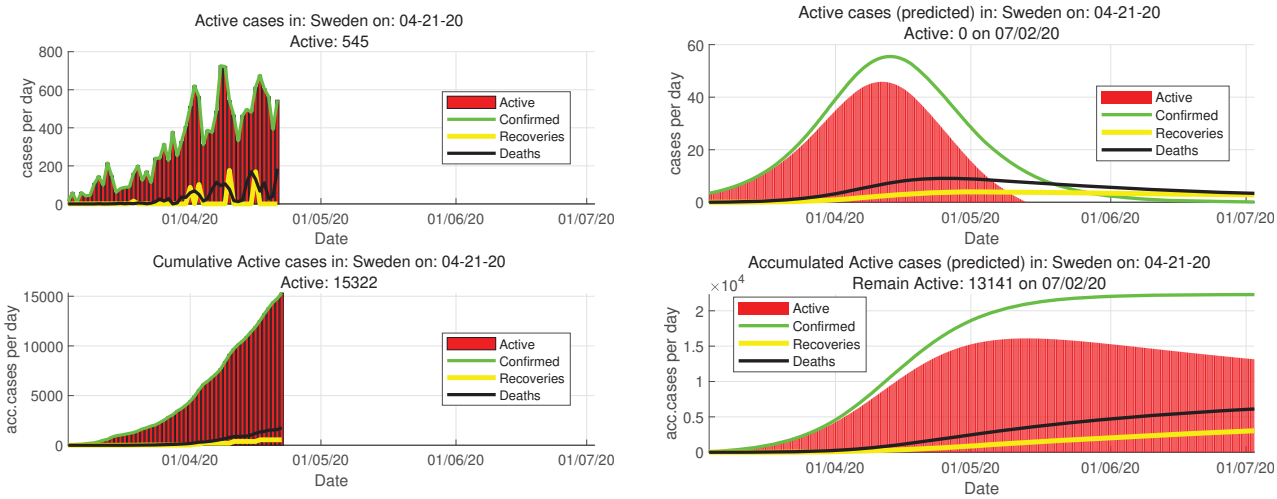

Fig. 58 Active cases reported and predicted from the SIR model without vital dynamics until the end of the first cycle of the activity of coronavirus for Sweden

point of the pandemic but are in the steady growth phase while Finland and Sweden passed the acceleration phase of the pandemic and are now on the way to enter the steady growth phase. Therefore, the peak of the disease is shown to have passed for all Nordic countries, but great caution as needed in the process of softening the prevention measures in the return to a normal life. Due to the negligence of some individuals, the coronavirus can exponentially increase the number of infected cases. So, hand washing and social distancing are very important, not just during the course of coronavirus but also after the ending phase and it will until a cure and vaccine can be found.

In the future, other epidemic models can be applied improving the results by applying the most accurate method, the so-called Bulirsch-Stoer. In addition, a combination of numerical methods can be applied for predictions, i.e. using the results of one method as initial conditions for the other numerical method.

VIII. Key Messages

- The bat is the source of the infection.
- China and Italy were identified as roots of the branching for most of the international epidemic.
- Italy and USA represent the bridge connecting the spread of the virus in Nordic countries.
- The prediction-corrector methods such as the ABM method are more accurate compared with other numerical methods under study and it traces signs of the coronavirus in the future after the first cycle of virus activity.
- All Nordic countries are in the deceleration phase of the pandemic.
- The secondary infections from one infected person are about one to two people and the average infectious period varies from 0 to 6 days.
- The first cycle of the pandemic is predicted to end in Denmark on 2 June 2020, in Norway on 11 May 2020, in Finland on 10 June 2020 and in Sweden on 14 June 2020.
- Preventative measures need to be rigorously respected until the cure and vaccine can be invented.
- The resurge of the infection will start in the middle July 2020 for Norway and Denmark, October 2020 for Sweden, and September for Finland.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Appendix A

Some predictions for other months of the year are illustrated in Figs. 59 and 60.

The activity of the virus will decrease on April but it never cease (for some time the activity will remain low) and the next cycle will start around middle of July 2020 for Norway and Denmark, October 2020 for Sweden, and September for Finland.

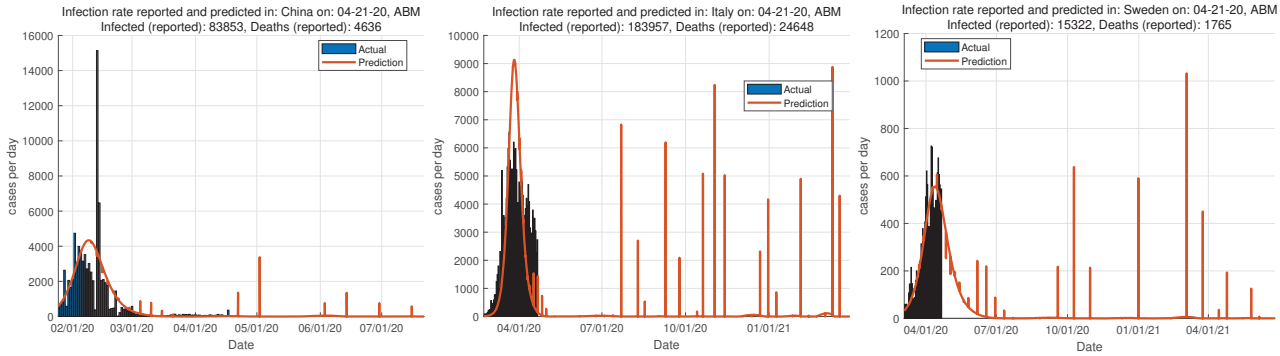

Fig. 59 The infection activity of SARS-CoV-2 predicted with SIR model without vital dynamics (ABM method) for China, Italy and Sweden

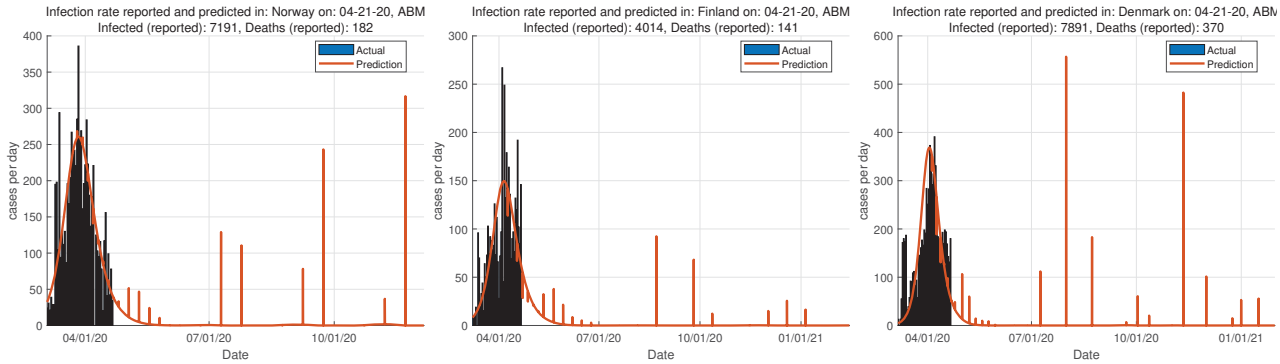

Fig. 60 The infection activity of SARS-CoV-2 predicted with SIR model without vital dynamics (ABM method) for Norway, Finland and Denmark

REFERENCES

[1] Bank, R. E., Coughran, W. M., Fichtner, W., Grosse, E. H., Rose, D. J., & Smith, R. K. *Transient simulation of silicon devices and circuits*. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems; 1985. **4**(4), 436–451.

[2] Boujakjian, H. Modeling the spread of Ebola with SEIR and optimal control. *SIAM Undergraduate Research Online*; 2016. 9, 299–310.

[3] Chen, Y., Cheng, J., Jiang, Y., & Liu, K. A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification. *Journal of Inverse and Ill-posed Problems*; 2020. **28**(2), 243–250.

[4] Cristianini, N., & Hahn, M. W. *Introduction to computational genomics: a case studies approach*. Cambridge University Press; 2006.

[5] Dormand, J. R., & Prince, P. J. A family of embedded Runge-Kutta formulae. *Journal of computational and applied mathematics*; 1980. **6**(1), 19–26.

[6] Hosea, M. E., & Shampine, L. F. *Analysis and implementation of TR-BDF2*. Applied Numerical Mathematics; 1996. **20**(1-2), 21–37.

[7] Guo, Z., Xiao, D., Li, D., Wang, X., Wang, Y., Yan, T., & Wang, Z. Predicting and evaluating the epidemic trend of ebola virus disease in the 2014-2015 outbreak and the effects of intervention measures. *PloS one*; 2016. *11*(4).

[8] Kermack, W. O., & McKendrick, A. G. A contribution to the mathematical theory of epidemics.*Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*; 1927. **115**(772), 700–721.

[9] Lai, D. Monitoring the SARS epidemic in China: a time series analysis.*Journal of Data Science*; 2005. **3**(3), 279-293.

[10] Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., ... & Fisman, D. Transmission dynamics and control of severe acute respiratory syndrome. *Science*; 2003. **300**(5627), 1966–1970.

[11] Loper, M. L. (Ed.). *Modeling and simulation in the systems engineering life cycle: core concepts and accompanying lectures*. Springer; 2015.

[12] Pandey, G., Chaudhary, P., Gupta, R., & Pal, S. SEIR and Regression Model based COVID-19 outbreak predictions in India. *arXiv preprint arXiv:2004.00958*; 2020.

[13] Poon, L. L. M., Chu, D. K. W., Chan, K. H., Wong, O. K., Ellis, T. M., Leung, Y. H. C., & Guan, Y. *Identification of a novel coronavirus in bats*. Journal of virology; 2005. **79**(4), 2001–2009.

[14] Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press; 2007.

[15] Rachah, A., & Torres, D. F. Predicting and controlling the Ebola infection. *Mathematical Methods in the Applied Sciences*; 2017. **40**(17), 6155–6164.

[16] Read, J. M., Bridgen, J. R., Cummings, D. A., Ho, A., & Jewell, C. P. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *MedRxiv*; 2020.

[17] Shampine, L. F. *Computer solution of ordinary differential equations: The Initial Value Problem*. W.H. Freeman, San Francisco; 1975.

[18] Shampine, L. F., & Reichelt, M. W. *The matlab ode suite*. SIAM journal on scientific computing; 1997. **18**(1), 1–22.

[19] Shampine, L. F., Reichelt, M. W., & Kierzenka, J. A. *Solving index-1 DAEs in MATLAB and Simulink*. SIAM review; 1999. **41**(3), 538–552.

[20] Tang, Z., Li, X., & Li, H. Prediction of New Coronavirus Infection Based on a Modified SEIR Model. *medRxiv*; 2020.

[21] https://www.ncbi.nlm.nih.gov/nuccore?LinkName=genome_nuccore_samespecies&from_uid=86693

[22] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/?gclid=EAIaIQobChMIptrK657D6QIVxeAYCh2I6g4YEAAYASACEgIHlv_D BwE

Gleda Kutrolli obtained her BA and MA in Mathematics and Computer Science at University of Tirana, Albania. For 6 years she has been a Quantitative Financial Analyst and a lecturer of Statistics and Actuarial Mathematics at University of Tirana, Albania. She is currently in her last year of her PhD studies in Mathematical Finance at University of Milano-Bicocca. Her research interests focus on: Statistics and Quantitative Financial Analysis: Statistical models, numerical techniques, optimization methods, pricing derivatives, risk measures and hedging strategies in financial world; Innovative Technology in Finance and Energy markets; Modelling and hedging strategies in commodity and energy markets.

Maksi Kutrolli obtained a Master of Science in Computer Science at University of Tirana, Albania. Currently he is working as Lecturer at Department of Computer Science, University of Aleksander Moisiu, Durres, Albania and is a co-founder of Smar-Tech. His research interests focus on: Machine Learning, Image Processing, Mobile Applications.

Etjon Meco has a Master of Science in Biochemistry. Currently working as a Research Associate in Genetox and Cell Services laboratories at TOXIKON CORP, Massachusetts, USA, for life science product testing. Performing various scientific tests and using cutting edge technology, his work helps pharmaceutical and medical devices industries to bring their products into market.