Predictive modeling of novel coronavirus spreading in European regions based on worldwide data-sets

Stanislav Vechet and Jiri Krejša

Abstract

Background: The novel coronavirus 2019-nCoV outbreak which was reported by World Health Organization (WHO) at 21th of January 2020 situation report spreads around the globe during next few weeks and causes serious health, social and financial issues. The primary epicentre was reported is Wuhan City, China and together with other affected regions in Asia remains the most monitored area. On the other hand, the European region was experiencing only a few cases until February 29th when the number of confirmed cases reached one thousand infections. Within a few days during the 8th of March the Italian most infected places were locked by restrictive quarantine. In this study we present probabilistic model to track the infection spread through the European regions based on the model adjusted with worldwide data-sets.

Methods: To track and predict the officially reported number of cases we developed probabilistic model with the ability to predict the number of confirmed cases one day ahead. The model has internal/hidden parameters which cannot be directly observed such as the number of infectious individuals or transmissibility rate which is represented by reproduction number $R_t$. The model starts with assumed number of infectious individuals and reproduction number $R_0$ and as more data are gathered from data-sets during time the internal parameters are further estimated. The model is updated each day as the new number of confirmed cases are reported. Particle filters algorithm is used as the back-end method due to its ability to handle multi-modal data distribution.

Results: Presented results show the performance of the probabilistic model which is able to handle short-term prediction (in number of days) of confirmed and recovered cases. The estimated reproduction rate is further used in long-term simulation which fits the data gathered world-wide. The one day prediction error is below 5% of nominal value and as we are located in Czechia the prediction model for our region was tested 4 days forward with the same error. The overall performance of the model was compared to data gathered from China due to the longer history of measurements.

Conclusion: We have proposed a probabilistic model which is used with particle filters to predict next moves of the confirmed cases. As a side effect we were able to model internal parameters as reproduction number or recovery rate which can be used for running long-term simulations of virus spreading.

Keywords: Novel coronavirus; Transmissibility; Probabilistic model; Predictive modeling; Particle filters
**Introduction**

Presented study focuses on modeling the growth of infections (number of confirmed cases) and transmissibility rate (represented by reproduction number $R_t$) of novel coronavirus 2019-nCoV (Covid-19) in European regions based on worldwide data-sets. The Covid-19 outbreak was reported by World Health Organization (WHO) on 21th January 2020 [1]. Another announcement was made by WHO on 30 January that the covid-19 already constituted a Public Health Emergency of International Concern [2]. After two months at the beginning of March the number of confirmed cases in Wuhan City, China reaches 67000 [3] infections and the European epicentre Italy was closed by restrictive quarantine with over 10000 cases on 11th March 2020 [3]. More European regions applied restrictive measures as closing public places, schools, universities etc.

In this study we present a probabilistic model to track the infection spread trough the European regions based on the model adjusted by worldwide data-sets.

Other approaches use mathematical models to simulate the dynamics of the infection spreading in population [4, 5, 6] or attempt to model the spread via commercial air travel [7, 8].

While our model is based on nonparametric implementation of the Bayes filter called Particle filters[9], there are method focused on other Bayes filters such as Kalman filters [10] or using Markov chain Monte Carlo methods [11].

**Materials and methods**

**Used data sets**

All reported experiments are based on online available data-sets [12] operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE)[13]. Data-sets which contain the total number of confirmed (reported) cases [3], recovered and deceased cases [14, 15] are used in simulation experiments for each Country/Region for each day starting from 22 January 2020 up to 22 March 2020.

**Simulation model**

The model estimates the Reproduction number $R$ which represents the transmissibility of virus changing in time as an average number of newly infected people generated by one infectious person. The initial constant $R_0$ is called the basic reproduction number[16] and there are researches focused on calculating the $R_0$[17, 18]. The actual average of infected cases generated from one person at time $t$ is called effective reproduction number $R_t$[6, 19] and is estimated by other researchers as $R_t = 6.47^{[20]}, R_t = 2.2^{[5]}, R_t = 2.68^{[11]}$.

The simulation model is based on prediction-correction mechanism [9] where in each prediction step the internal parameters estimation is made and during correction step the model parameters are updated with measurement (if available). The internal state of the model $x_t = \begin{pmatrix} I_t & C_t & Re_t & R_t \end{pmatrix}^T$ is represented by effective reproduction number $R_t$, number of infectious individuals $I_t$ which are not detected, number of confirmed cases $C_t$ and number of recovered cases $Re_t$ which can be obtained from public reports. The measurement taken in time $t$ is represented by variable $z_t$. 
The algorithm uses particle filters for representing posterior distribution \( p(x_t|z_t) \) over possible states \( x_t \) based on measurements \( z_t \) by set of particles \( X_t \). The simulation time step was chosen as one day due to the nature of used data-sets. The input for the model (measurement) is number of confirmed cases for each day. Due to the fact that the available data set provides the information about confirmed cases, the model is used to predict the confirmed cases \( C \) based on estimated effective reproduction number \( R_t \) and number of infectious individuals \( I_t \).

One iteration of proposed model is given by following steps:

1. \( \bar{X}_t = X_t = \emptyset \)
2. \( x_t = \begin{pmatrix} I_t \\ C_t \\ R_t \\ \end{pmatrix} \)
3. \( p(x_t|x_{t-1}^{[n]}) \sim \begin{pmatrix} \frac{R_t}{R_{t-1}} \sim N \left( \frac{R_t - 1}{R_{t-1}}, \sigma^2 \right) \\ I_t = I_{t-1} + \frac{1}{\text{incubation period}} - I_{t-1}N \left( \frac{1}{\text{incubation period}}, \sigma^2 \right) \\ C_t = C_{t-1} + \frac{1}{\text{recovery period}} - C_{t-1}N \left( \frac{1}{\text{recovery period}}, \sigma^2 \right) \\ R_t = R_{t-1} + C_tN \left( \frac{1}{\text{recovery period}}, \sigma^2 \right) \end{pmatrix} \)
4. \( x_{t}^{[n]} \sim p(x_t|x_{t-1}^{[n]}) \)
5. \( w_{t}^{[n]} = p(z_t|x_{t}^{[n]}) \sim \frac{1}{\sigma \sqrt{2\pi}} e^{-0.5 \left( \frac{(z_t-x_{t}^{[n]})}{\sigma^2} \right)^2} \)
6. \( \bar{X}_t = \bar{X}_t + \langle x_{t}^{[n]}, w_{t}^{[n]} \rangle \)
7. draw \( i \) with probability \( \propto w_{t}^{[i]} \)
8. add \( x_{t}^{[i]} \) to \( X_t \)

Step 1 and 2 represents the initialization of particles and state vectors. Step 3 and 4 describes the sampling of new set of particles from the original distribution of samples in previous time step \( t-1 \). For each particle \( x_t \) weight \( w_t \) is calculated based on measurement \( z_t \) and the weighted set of samples \( \bar{X}_t \) is updated. Finally the new set of particles \( X_t \) is created in steps 7 and 8, where the new states are resampled with probability given by its weights.

The parameter \text{incubation period} was set to 5.2 days according to WHO report[21], work presented by Li et al.[5] and calculations based on disease onset to diagnosis [22]. Parameter \text{recovery period} was originaly set to 10 days according to parameteres used different models e.g. epidemic SEIR models [23, 24] where it represents the time form diagnosis to total recovery.

**Results**

Simulation experiments were performed on regions located in China where the proposed model was tested and then applied to regions located within Europe. While many regions in Europe reach the similar number of infectious individuals we show on Fig. 1 the comparison of China - Hubei with Europe regions which reaches the number of confirmed cases more than one thousand individuals by the 22 March 2020.

Outputs from described model for China-Hubei are shown of figure 1, where the model outputs are numbers of confirmed cases, recovered cases and infectious individuals. For each day the prediction from proposed model was taken and afterwards
compared to the number of confirmed (reported) cases for particular day, after complete set of measurements for all evaluated days (from 23 January 2020 to 22 March 2020) the $RMSE = 2516.6$ and average $R_t = 1.09$ were calculated. From the simulation the basic reproduction number $R_0 = 1.4$ and number of infectious individuals $I_0 = 998$ were estimated. We can compare the modeled data with number of confirmed and reacovered cases to see the performance of the model. The number of infectious individuals is only informative value because we cannot compare it with real data. There is also an approach to get informative number of real infections in the population presented by Wu and McGoogan [25], however is impossible to get similar data-sets for european locations.

The model itself can be used for simulation of estimated long-term cases. For this purpose the model is initialized with the confirmed cases $C_0 = 444$ as reported 23th of Jan 2020, the number of infectious individuals estimated by particle filters $I_0$ and number of recovered cases $Re_0 = 0$. Long-term simulation is shown on figure 3 and the results are comparable with real-world scenario.

The one day predictions for European regions are represented by Czechia as our home country (Fig. 4), Italy (Fig. 5) as the highest number of confirmed cases by the 22th of Marh 2020 and Germany (Fig. 6) as the region with the first confirmed european case.

The model is capable of short-term prediction for more than one day ahead. The result for four days prediction is shown on figure 7 where the predicted number of cases for Wed 25 Mar 2020 was 1765 and for Thu 26 Mar 2020 the prediction was 1796 cases, while the real reported number of confirmed cases by the Ministry of Health of the Czech Republic [26] was 1763 cases for Wed resp. 2022 for Thu.
Discussion

Even there are data-sets which cover the global infection spreading, there could be problem with the data gathered for prediction analysis. We have found the differences in data presented in used data-sets [3] and data tracked by our team and presented locally in Czechia (Tab. 1). While epidemic in Czechia is actually at the very beginning it is difficult to track the confirmed cases and update the model. The estimated reproduction number average for Czechia $R_t = 1.3$ is in correlation with the growing number of confirmed cases and also with the restrictive measures applied locally to high schools and universities which should lead to infection elimination.

| Date          | Worldwide data-sets for Czechia | Czechia local data sources |
|---------------|---------------------------------|-----------------------------|
| Wed 2020-03-04| 8                               | 5                           |
| Thu 2020-03-05| 12                              | 8                           |
| Fri 2020-03-06| 18                              | 19                          |
| Sat 2020-03-07| 19                              | 26                          |
| Sun 2020-03-08| 31                              | 32                          |
| Mon 2020-03-09| 31                              | 38                          |
| Tue 2020-03-10| 41                              | 63                          |
| Wed 2020-03-11| 91                              | 94                          |

All presented results show that the prediction error for number of confirmed cases is below 5% of nominal value which exceeds our expectations. There is also a drawback in the usage of reported number of recovered cases which are only used for RMSE calculation not for the model update. Even if this value is not used in the correction step of the particle filter algorithm, the model still tracks the real values of the recovered cases as shown of figure 2.

The estimated average reproduction numbers for representative locations around Europe $R_t = 1.28$ for Italy, $R_t = 1.31$ Czechia and $R_t = 1.22$ Germany together with the estimated $R_t = 1.1$ for China-Hubei are lower than the reproduction numbers estimated by other researches at the Covid-19 outbreak $R_t = 6.47$[20] for Wuhan during January 2020 which can be due to different history data analysis. On the other hand, the lower estimated number e.g. for China-Hubei after almost two months shows that the infection spreading slows down. For the next weeks we can expect that the applied restrictive measures in Europe will similarly help to get the infection under control although high requirements for the health infrastructure are expected and it can take a few weeks to reduce the reproduction number.

Conclusion

Proposed probabilistic model shows the capability to handle short-term predictions for various regions in Europe. Moreover, from the simulation experiments it seems
to be reasonable to use such model for overall trends analysis. While the novel coronavirus epidemic is still evolving, as more data will be available the model parameters can be adjusted more precisely and comparison with other classical models e.g. SEIR can be performed.

Availability of data and materials
The datasets supporting the result are available in public repositories as referenced.

Abbreviations
 Covid-19: Coronavirus disease 2019
 2019-nCoV: 2019 novel coronavirus
 WHO: World Health Organization
 RMSE: Root Mean Square Error
 SEIR: Susceptible-Exposed-Infectious-Removed

Acknowledgements
We thank to Brno University of Technology for the support during the novel coronavirus epidemic.

Funding
This work is support by Institute of Solid Mechanics, Mechatronics and Biomechanics, Brno University of Technology, Czech Republic (VAV13250).

Author information
Affiliations
Brno University of Technology, Institute of Solid Mechanics, Mechatronics and Biomechanics, Technicka 2896/2, 616 69, Brno, Czech Republic
Stasílavl Vechet, Jiří Krejsa

Contributions
S.V. had the idea to predict the confirmed cases by Particle Filters and have created and tested the model. J.K. contributed to conceptual ideas and helps to writing the article. All authors have read and approved the published version of the manuscript.

Ethics declarations
Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1 Brno University of Technology, Institute of Solid Mechanics, Mechatronics and Biomechanics, Technicka 2896/2, 616 69 Brno, Czech Republic. 2 Brno University of Technology, Institute of Solid Mechanics, Mechatronics and Biomechanics, Technicka 2896/2, 616 69 Brno, Czech Republic.

References
1. Novel Coronavirus(2019-nCoV) SITUATION REPORT-1, 21 JANUARY 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4
2. WHO. Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV). https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)
3. Number of Confirmed Cases. https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Confirmed.csv
4. Chen, T.-M., Rui, J., Wang, Q.-P., Zhao, Z.-Y., Cui, J.-A., Yin, L.: A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Infectious Diseases of Poverty 9(1), 24 (2020). doi:10.1186/s40249-020-00640-3
5. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T.Y., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z.: Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. New England Journal of Medicine 0(0), (0). doi:10.1056/NEJMoa2001316. https://doi.org/10.1056/NEJMoa2001316
6. Wang, H., Wang, Z., Dong, Y., Chang, R., Xu, C., Yu, X., Zhang, S., Tsamlag, L., Shang, M., Huang, J., Wang, Y., Xu, G., Shen, T., Zhang, X., Cai, Y.: Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. Cell Discovery 6(1), 10 (2020). doi:10.1038/s41421-020-0148-0

7. Bogoch, I.I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M.U.G., Khan, K.: Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. Journal of Travel Medicine (2020). doi:10.1093/jtm/taaa008. taaa008.

8. How Computer Scientists Are Trying to Predict the Coronavirus’s Next Moves. https://spectrum.ieee.org/the-human-os/biomedical/devices/predicting-the-coronavirus-next-moves

9. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics (Intelligent Robotics and Autonomous Agents). The MIT Press, ???. (2005)

10. Kremer, R.: Using Kalman Filter to Predict Corona Virus Spread. https://towardsdatascience.com/using-kalman-filter-to-predict-corona-virus-spread-72d91b74c8

11. Wu, J.T., Leung, K., Leung, G.M.: Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in Wuhan, China: a modelling study. The Lancet 395(10225), 689–697 (2020). doi:10.1016/S0140-6736(20)30260-9

12. Novel Coronavirus (COVID-19) Cases, Provided by JHU CSSE. https://github.com/CSSEGISandData/COVID-19

13. Johns Hopkins University, Center for Systems Science and Engineering (CSSE) at JHU. https://systems.jhu.edu/research/public-health/ncov/

14. Number of Recovered Cases. https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Recovered.csv

15. Number of Deceased Cases. https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Deaths.csv

16. Imai, N. et Al. Report 3: Transmissibility of 2019-nCoV, 2020. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news–wuhan-coronavirus/

17. Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., Wang, M.H.: Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. International Journal of Infectious Diseases 92, 214–217 (2020). doi:10.1016/j.ijid.2020.01.050

18. Zhao, S., Musa, S.S., Lin, Q., Ran, J., Yang, G., Wang, W., Lou, Y., Yang, L., Gao, D., He, D., Wang, M.H.: Estimating the unreported number of novel coronavirus (2019-ncov) cases in China in the first half of January 2020: A data-driven modelling analysis of the early outbreak. Journal of Clinical Medicine 9(2) (2020). doi:10.3390/jcm9020388

19. Haydon D. T., S.D.J.M.L.F.J.K.W.J. Chase–Topping M., J., W.M.E.: The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. 270. Proc. R. Soc. Lond. B. doi:10.1098/rspb.2002.2191

20. Tang, S., Wang, Z., Li, M., Bragazzi, N., Xiao, Yu, D.: Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. Journal of Clinical Medicine 9, 462 (2020). doi:10.3390/jcm9020462

21. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/docs/default-source/coronaviruses/who-china-joint-mission-on-covid-19-final-report.pdf

22. Yang, Y., Lu, Q., Liu, M., Wang, Y., Zhang, A., Alali, N., Dean, N., Longini, I., Halloran, M.E., Xu, B., Zhang, X., Wang, L., Liu, W., Fang, L.: Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. medRxiv (2020). doi:10.1101/2020.02.21.20021675. https://www.medrxiv.org/content/early/2020/02/21/2020.02.20021675.full.pdf

23. Chowell Gerardo, N.H., M.A, B.L.: Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. 43. R. Soc. Interface. doi:10.1098/rspi.2002.2191

24. Chowell, G., Hengartner, N.W., Castillo-Chavez, C., Fenimore, P.W., Hyman, J.M.: The basic reproductive number of ebola and the effects of public health measures: the cases of congo and uganda. Journal of Theoretical Biology 229(1), 119–126 (2004). doi:10.1016/j.jtbi.2004.03.006

25. Wu, Z., McGoogan, J.M.: Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in China: Summary of a report of 72 314 cases from the chinese center for disease control and prevention. JAMA 2020. doi:10.1001/jama.2020.2649

26. COVID-19: Actual Numbers of Coronavirus Diseases in the Czech Republic, Ministry of Health of the Czech Republic. https://onemocneni-aktualne.mzcr.cz/covid-19