Proceedings of CONF-APMM 2026 Symposium: Applied Mathematics and Quantitative Modeling
DOI: 10.54254/2753-8818/2026.PU36314

SEIR-Based Modeling of COVID-19: Comparative Evaluation of Parameter and Equation Choices

Shuling Meng

Tsinglan School, Dongguan, China
Eric.Meng_27@tsinglan.org

Abstract. This article compares the accuracy of susceptible-exposed-infected-recovered (SEIR) model and two improved models in terms of case number fitting and peak prediction during the early stage of the epidemic in the same province (from January 2020 to March 2020), examining whether there are any issues with the models. The paper uses some models and data as examples to analyze the advantages and disadvantages of different models. That's an important task because models may not fit reality satisfactorily or need more time to calculate if the models are not good enough. In some cases, researchers even need to give up the models. In the research, the author tidied up some papers that told people how to use SEIR models and their variants to analyze coronavirus disease 2019 (COVID-19), and the author found that in most cases, people need to decrease the number of parameters and use the real parameters. There are also a few exceptions. The scholars employed some special algorithms or other mathematical tools to enhance the accuracy of the model and effectively avoid a series of problems caused by excessive parameters and variables. However, according to the author's observation, the number of such articles is relatively small and requires a very high level of mathematical knowledge. Most articles still use fewer variables and employ actual parameters.

Keywords: SIR, SEIR, epidemic dynamics

1. Introduction

Infectious diseases pose a huge threat to the real world. Countless scholars have made tremendous efforts to control them and have achieved many encouraging results, including the establishment of the discipline of infectious disease dynamics and numerous subsequent models. Infectious disease dynamics is an important method for theoretical quantitative research on infectious diseases, playing a significant role in understanding the transmission patterns and development trends of infectious diseases. The transmission of infectious diseases is influenced by many factors, such as the incubation period, the latent period, the recovery rate, and the transmission rate. When a new infectious disease emerges, reducing the disease transmission rate is the most direct and effective control measure. Through communication networks, people can quickly obtain information about the disease and establish disease prevention awareness, reducing contact and lowering the transmission rate. The deterministic model of infectious diseases was initially proposed by Kermack

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.PU36314

and McKendrick in 1927 and is known as the susceptible-infected-recovered (SIR) model, laying the foundation for the research of infectious disease dynamics.

In practical problems, the number of the population always changes due to external and internal disturbances. To make the model more realistic, this article will select several articles to analyze the models the researchers have chosen, as well as their advantages and disadvantages, and make some improvements at the end. These include "Construction and Research of an Infectious Disease Model Based on the Transmission Characteristics of the coronavirus disease 2019 (COVID-19), "Application of the Modified the susceptible-exposed-infected-recovered (SEIR) model Infectious Disease Dynamics Model in the Prediction and Evaluation of COVID-19 Epidemic in Hubei Province", etc.

The SIR is the simplest model, which divides the people with the infectious disease into three parts: susceptible population, infected population, and recovered population. To fit the real situation, subsequent scholars have added variables such as the latent population, the deceased, and parameters such as isolation factors and vaccine factors in the transmission process. In addition, some people also use other mathematical tools, such as constructing Lyapunov functions and using Itˆo formulas, applying LaSalle invariance principles, etc., to reduce fitting errors. Although these things make the model more accurate, those things may also cause the model to overfit, the equation system to be too rigid, or difficult to solve. Even worse, sometimes these formulas may not be able to well cover or describe time-limited situations. This is the reason why this article discusses analyzing the use of models and formulas.

2. Literature review

The predecessors have made many contributions to people. On the basis of the most classic SEIR (susceptible, exposed, infected, recovered) model, Cao Shengli, Feng Peihua, Shi Pengpeng, and others modified it by adding four parameters: isolation ratio, infection probability, contact rate, and effective contact coefficient. The original four equations were expanded as follows [1]:

$$dS / dt = -[\rho c\beta + \rho cq(1-\beta)]S(I+\theta E) + \lambda S_q \tag{1}$$

$$dE / dt = \rho c\beta(1-q)S(I+\theta E) - \sigma E \tag{2}$$

$$dI / dt = \sigma E - (\delta_I + \alpha + \gamma_I)I \tag{3}$$

$$dS_q / dt = \rho cq(1-\beta)S(I+\theta E) - \lambda S_q \tag{4}$$

$$dE_q / dt = \rho c\beta qS(I+\theta E) - \delta_q E_q \tag{5}$$

$$dR / dt = \gamma_I I + \gamma_H H \tag{6}$$

$$dH / dt = \delta_I I + \delta_q E_q - (\alpha + \gamma_H)H \tag{7}$$

Using this improved model, the current model's improvement is in good agreement with the actual data of the epidemic situation in Hubei from January 23 to February 12, 2020 [1]. However, there are still differences between the model and reality. The predicted peak of current infected cases has a one to two-day difference from reality, and the number of equation systems is too large, which is objectively unfavorable for calculation [1]. At the same time, Zhu Maichang, Bin Zhen, Sun

DOI: 10.54254/2753-8818/2026.PU36314

Zengxin, and others have brought us a new, improved model [2]. Those people improved the SEIR model to the susceptible-exposed-asymptomatic-infected-hospitalized-recovered (SEAIHR) model , with the principle of adding hospitalized isolators, latent asymptomatic carriers, latent recoverers, and deceased patients. Those people rewrote the original equations as:

$$dS_t/dt = - (E_t\lambda_1 + A_t\lambda_2) \tag{8}$$

$$dE_t/dt = (1 - m) * (E_t\lambda_1 + A\lambda_2) - E_t\beta_1 - E_t\mu \tag{9}$$

$$dA_t/dt = m(E_t\lambda_1 + A_t\lambda_2) - A_t\beta_3 - A_t\beta_2 \tag{10}$$

$$dIt/dt = E\mu - I\rho \tag{11}$$

$$dH_t/dt = -E_t\beta_1 + I\rho H(P_r + P_d) \tag{12}$$

$$dC_t/dt = A_t\,\beta_3 \tag{13}$$

$$dR_t/dt = HP_r \tag{14}$$

$$dD_t/dt = HP_d \tag{15}$$

The model has achieved good results in fitting the daily mortality rate, but there is still a discrepancy between the daily cure rate and the actual situation [2]. Moreover, the large number of equations makes the calculation process cumbersome. Compared with the previous model, this model is more inclined to calculate the number of people who are cured and die each day, rather than the number of people who are infected.

On the other hand, some people are more inclined to apply new mathematical methods to the study of infectious diseases rather than modifying the existing models. For instance, Zhao Yanjun, Su Li, Sun Xiaohui, Li Wenxuan, and others applied logistic growth, Lyapunov functions, Itô formulas, and LaSalle invariance principles to infectious disease models, thereby describing the persistence of epidemic systems in reality [3].

3. Methodology

The SIR model, based on the SI model, has some improvements. It adds a recovered group R. The total population remains constant at N. The R assumption in this model is that patients have permanent immunity, so the deceased population is also classified as R. Parameter α represents the average time for recovered individuals to acquire immunity. In the SIR model, there are only two independent dynamic variables, I and S, and their phase trajectory is: If the studied infectious disease has a certain incubation period, then it will be a good idea to use the SEIR model. E represents the population with an incubation period. Its total population remains constant. This model has equilibrium points at S = N and S = γ/β. The former indicates that the disease is eliminated, and the latter indicates that the disease is in an epidemic state [4]. From this, it can be seen that although these basic models have fewer variables and parameters, the variables and parameters do not quite conform to the actual pathological conditions. Therefore, Cao Shengli et al. modified the SEIR model, attempting to make it more in line with the actual situation. In fact, it should be acknowledged that this model is quite good.

DOI: 10.54254/2753-8818/2026.PU36314

In Figure 1, it can be seen that this model fits well with the actual data. Before February 12th, it closely matched the actual data. In Figure 2, due to its high degree of fit with reality, it also has considerable credibility for predictions that do not take into account the incubation period. The best result is shown in Figure 3, which precisely and perfectly predicted the critical stage of the infection peak. However, the models also have some problems. For example, the setting of initial parameters, such as the number of infected individuals, was multiplied by 1.5 in the official data, which was explained as considering the number of undetected individuals [1]. However, the value is probably incorrect. As of January 25, 2020, 23:59, the real-time dynamic data of the 2019-nCoV epidemic situation reported by People's Daily was: 1,408 confirmed cases and 2,032 suspected cases [5, 6]. The data provided by the World Health Organization shows that a total of 1,320 confirmed cases have been reported for novel coronavirus (2019-nCoV) globally. Of the 1,320 cases reported, 1,297 cases were reported from China, including Hong Kong SAR (5 confirmed cases), Macau SAR (2 confirmed cases), and Taipei (3 confirmed cases) [7]. The data recorded by Baidu Encyclopedia shows that as of January 25th, a total of 1,287 cases had been confirmed, and 41 deaths occurred in 29 provinces (autonomous regions and municipalities). 38 patients had been discharged. 1965 suspected cases had been reported in 20 provinces (autonomous regions and municipalities) [8]. Based on the actual control situation, there is a certain proportion of symptomatic infected individuals who have not been discovered. The original text states that the fitting degree was relatively good from January 23 to February 11, 2020. So, only two days later, the number of infections could increase by 6 times. This speed is not realistic because subsequent models and other models did not reflect this result, even without considering the control factors [9-11]. Therefore, the initial value of the number of infected individuals in this article is overly optimistic. And using overly optimistic data, the model fits the reality better, so the model may have problems. In conclusion, the model may be correct in judging the general trend of the epidemic, but the specific values have problems. On the other hand, Zhu Maochang et al. brought SEAIHR model , which has fewer parameters and less computational complexity (with a total of ten parameters, some of which need to be inferred) compared to the SEIR model. The lines have a very close fit to the first-stage active cases (infected individuals) in Wuhan, with an average relative error value 34.4% smaller than that of SEIR [2]. The drawback is that, possibly due to the reasons for data collection, the fitting of the daily cure rate and the subsequent new epidemic situation in Xinjiang is not particularly good. It is also not excluded that the model has poor fitting for the cure rate [2].

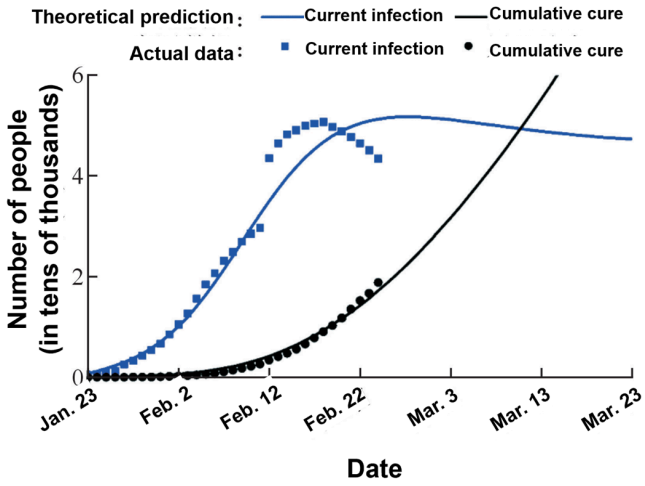

Figure 1. Comparison of the theoretical estimates of the number of infected and recovered by the modified SEIR model with and without latent infectivity [1]

DOI: 10.54254/2753-8818/2026.PU36314

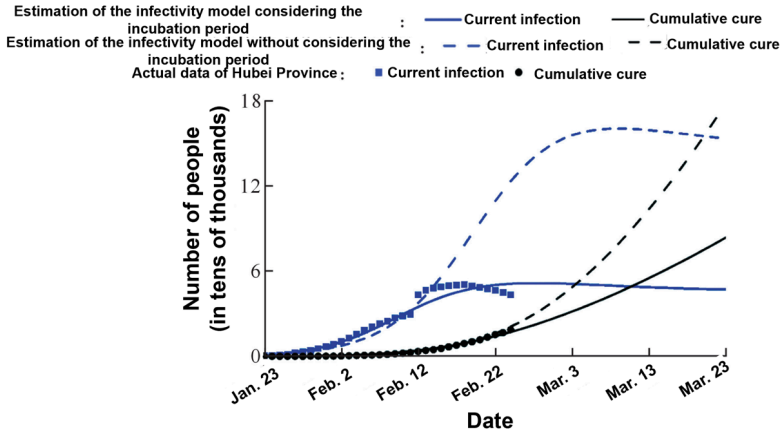

Figure 2. Prediction of the travel of the COVID-19 epidemic in Hubei province by the modified SEIR model [1]

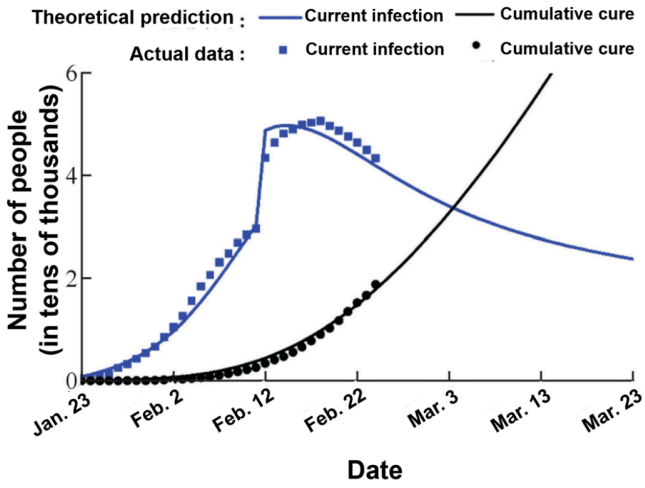

Figure 3. Forecast of COVID-19 epidemic trend in Hubei province after the SEIR model was revised again according to the modification of clinical diagnosis standard [1]

4. Results

So, to conclude those papers, most of the time, the more parameters, the more accurate the results. It will fit better with the real cases. From the results of the aforementioned paper, it can be seen that the model with the increased parameters fits better than the original model. However, the more parameters there are, the harder it is to calculate or to let the machine learn. For example, there are only four equations in SEIR, but for SEIAHR, there are six equations. What's more, when scholars try to analyze the models or just the propagation process, they would like to find the part that increases or decreases very gently in the function. Then, scholars will find the derivative of the function. Because this part changes very gently, the derivative can be used for a bit more time. However, more equations mean more functions, so it will probably be hard to find a part where every function changes gently. So it will be hard to analyze. In some other cases, scholars aim to discover some general patterns of the spread of the plague. However, due to the large number of parameters, scholars will find it difficult to determine which parameter or which parameters are decisive in the spread of the plague.

DOI: 10.54254/2753-8818/2026.PU36314

5. Discussion

From the above materials, it can be concluded that, in the author's opinion, for a good model, the output results should always select the appropriate parameters. All the parameters should be based on valid reasoning and evidence, and it would be best to combine relevant data. Furthermore, if possible, the number of parameters should be minimized as much as possible. This not only simplifies the calculation process but also avoids the possibility of the model sacrificing overall performance for minor improvements. However, if increasing the number of parameters can make the model more in line with reality, it would also be an option. This paper merely summarizes the advantages and disadvantages of previous studies based on the author's own perspective. It is hoped that this can serve as a reference for future scholars. However, there are some issues that this article has not addressed. So, how many equations and how many parameters are appropriate for a model? Could the increase in equations and parameters be related to the computational difficulty of the model? From a pathological perspective, how many equations and parameters are appropriate? How to make a trade-off between model authenticity and computational complexity? The author of this article believes that subsequent scholars can take this as a research direction.

6. Conclusion

As mentioned above, this article merely summarizes the gains and losses of some of the papers that the author has recently read. It specifically summarizes the issues, including improper selection of parameters when presenting the results, as well as some problems with parameter settings. Regarding the selection of specific data, the author of this article believes that it is very important to extensively collect first-hand information and calculate the specific parameters from it. One must absolutely not make estimations on one's own. When it is impossible to calculate certain parameters based on the existing first-hand data, one should instead refer to the parameters of other infectious diseases or calculate them based on pathological data, rather than making an estimate on one's own. The main point of this article's author is that, in most cases, the parameters and equations in the model should be kept to a minimum in order to simplify the calculation. The simplicity of the calculation should be taken into account before considering the accuracy of the model, unless the computing power can be guaranteed. It is hoped that this can serve as a hint for the establishment of subsequent models. Other scholars might be able to build upon the content of this article to further explore how to establish a good model to refute the conclusions presented here and to create their own system, For instance, people can identify certain equations to calculate the ratio of the computational difficulty of the equations and parameters involved to the accuracy of the model for a specific disease and model, thereby determining the optimal number of equations and parameters for that model or disease.

References

[1] Cao, S., Feng, P., & Shi, P. (2020). Study on the epidemic development of COVID-19 in Hubei Province by a modified SEIR model. *Journal of Zhejiang University (Medical Sciences), 49*(2), 178–184. https://doi.org/10.3785/j.issn.1008-9292.2020.02.05.

[2] Zhu, M., Sheng, B., & Sun, G. (2023). Construction and research of infectious disease models based on the transmission characteristics of COVID-19. *Complex Systems and Complexity Science, 20*(2), 29–37. https://doi.org/10.13306/j.1672-3813.2023.02.004.

[3] Zhao, Y., Su, L., Sun, X., & Li, W. (2024). Qualitative analysis of a stochastic SIRS model with logistic growth and psychological effects. *Journal of Engineering Mathematics (China), 41*(3), 469–480. (Chinese).

[4] Baidu Baike. (2025, July). Infectious disease model (in Chinese). *Baidu Baike*.

DOI: 10.54254/2753-8818/2026.PU36314

[5] Zhou, T., Liu, Q., Yang, Z., Liao, J., Yang, K., Bai, W., Lü, X., & Zhang, W. (2020). Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *Journal of Evidence-Based Medicine, 13*(1), 3–7. https://doi.org/10.1111/jebm.12376.

[6] Yan, Y., Chen, Y., Liu, K., Luo, X., Xu, B., Jiang, Y., & Cheng, J. (2020). Modeling and prediction for the trend of outbreak of NCP based on a time-delay dynamic system (in Chinese). *Scientia Sinica Mathematica, 50*, 1–8. https://doi.org/10.1360/SSM-2020-0026.

[7] World Health Organization. (2020, January 25). *Novel coronavirus (2019-nCoV) situation report—5*. *World Health Organization*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200125-sitrep-5-2019-ncov.pdf

[8] Baidu Baike. (2020). *2020 China COVID-19 pandemic chronology* (in Chinese). *Baidu Baike*. https://baike.baidu.com/

[9] Gan, Y., Wu, Y., & Wang, J. (2021). Epidemic trend prediction model of COVID-19. *CAAI Transactions on Intelligent Systems, 16*(3), 528–536. https://doi.org/10.11992/tis.202008037

[10] He, S., Peng, Y., & Sun, K. (2020). SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynamics, 101*(3), 1667–1680. https://doi.org/10.1007/s11071-020-05743-y

[11] Gupta, R., Pandey, G., Chaudhary, P., & Pal, S. (2020). SEIR and regression model-based COVID-19 outbreak predictions in India. *medRxiv*. https://doi.org/10.1101/2020.04.01.20049825