Parametrization of bass diffusion model on COVID-19 first wave data

A Škraba¹, B Vavtar¹, V Stanovov², E Semenkin³ and R Stojanović³

¹ University of Maribor, Cybernetics & Decision Support Systems Laboratory
Faculty of Organizational Sciences, Kidričeva cesta 55a, 4000 Kranj, Slovenia
² Reshetnev Siberian State University of Science and Technology Institute of
Computer Science and Telecommunications
31 Krasnoyarskiy Rabochiy Ave., Krasnoyarsk, 660037, Russian Federation
³ University of Montenegro Faculty of Electrical Engineering Džordža Vašingtona bb.,
Podgorica, Montenegro

E-mail: andrej.skraba@um.si

Abstract. The paper addresses the development of the system for the parameterization of the
Bass Diffusion Model on Covid-19 Data. Continuous model developed by the principle of
system dynamics is presented. The parameters of the model are tuned to fit the data with genetic
algorithm in Powersim simulation tool. The validation is performed by standard simulation
validation techniques. The input data used in the experiments were gathered by publically
available databases. In particular, the data collected in Austria, France, Italy, South Korea,
Slovenia and Swiss were considered in this study. The results have shown that the simple model
with is capable of predicting the growth with S-shaped curve with high accuracy.

1. Introduction
The first wave of the COVID-19 pandemic in 2020, between February and June in several cases
exemplifies S-shaped growth, according to the publicly available databases [1, 2]. In numerous
countries, the S-shaped growth of the number of infected is evident however, one could observe in some
cases rather irregular growth, which might be a consequence of counter measures or errors in data. As
it has been indicated, the relatively simple SEIR model could be used to accurately fit the real case
trajectory despite the complexity of considered system [4].

In order to estimate the appropriateness of using the basic Bass diffusion model [3, 11] to model
pandemic dynamics, the mathematical model will be realized by the principles of System Dynamics
[11], calibrated and validated which might be used for future estimates in similar pandemic cases.

2. Methodology
The Bass model is the basic structure, that is present in more advanced SEIR models and as such, this
model has been the main tool in the field of diffusion research [6, 7, 8, 9]. The model could capture a
different diffusion patterns observed in the cases of pandemic as well as in other processes such as
marketing and information spread. The Bass diffusion model is defined as [3, 11]:
\[ I(t) = I(0) + \int_0^t R(t) \, ds \]

\[ S(t) = S(0) - \int_0^t R(t) \, ds \]

where \( I(t) \) represents the number of infected in time with units [people], \( I(0) \) initial number of infected, \( R(t) \) represents the newly infected with units [people/day], \( S(t) \) represents the number of susceptible over time. Rate of infected \( R(t) \) decreases the number of susceptible and on the other hand increases the number of infected. The Rate of infected \( R(t) \) is defined as:

\[ R(t) = c \lambda \rho I(t) \]

where \( c \) is the concentration of susceptible people, \( \lambda \) is social factor, \( \rho \) infection spread factor.

Concentration \( c \) is computed as the ratio between number of susceptible and sum of susceptible and infected:

\[ c = \frac{S(t)}{S(t) + I(t)} \]

Defined model provides an insight into the dynamics and key relationships between variables. In order to estimate the value of Bass diffusion model, the results of the simulations will be compared with the real data. Data for validation [1, 2] were taken from 12.3.2020 – 25.5.2020 for Slovenia, 1.2.2020 – 4.5.2020 for South Korea, 26.2. 2020 - 22.5.2020 for Swiss, 21.2.2020 - 8.6.2020 for Italy, 27.2.2020 - 28.5.2020 for France and 26.2.2020 - 3.5.2020 for Austria.

The model will be realized with Powersim simulation tool by the principles of system dynamics. Simulation method used was Runge-Kutta 4 with fixed step 0.125. In order to perform the calibration of the model the genetic algorithm solver will be used with criteria function of minimum sum of squared errors.

Figure 1 shows the adapted Bass-diffusion model build by the principles of System Dynamics. There are two stocks for number of Susceptible and number of Infected. The concentration of susceptible decreases over time since the stock of the susceptible depletes and on the other hand, this increases the number in Infected stock. Interaction between susceptible and Infected is modelled by the multiplication considering the social factor and Infection spread factor. This two factors determine the intensity of the infection spread. In order to calibrate the model, the Sum of Squared Error is realized by integrating the error between Real Data and simulation data of Infected. The integral is represented by the SSE stock.

![Simulation model build by the principles of System Dynamics.](image)
Figure 2 shows Bass diffusion model adapted for epidemics spread. The numbers are taken for the case of Slovenia. By entering the adequate numbers, the simulation could be performed for other countries although, the condition of S growth should be meet in order to use the model.

init  Infected = 49
flow  Infected = +dt*InfectionRate
doc  Infected = Number of infected.
init  SSE = 0
flow  SSE = +dt*ErroSquared
doc  SSE = Sum of squared errors. Criteria function. This value should be as small as possible.
init  Susceptible = InitNumber
flow  Susceptible = -dt*InfectionRate
doc  Susceptible = Potential number of susceptible.
aux ErroSquared = Error^2
doc  ErroSquared = Squared error.
aux  InfectionRate = InfectionSpreadFactor*Contacts
doc  InfectionRate = Number of newly infected.
aux  Concentration = Susceptible/S_and_I
doc  Concentration = Concentration of susceptible.
aux  Contacts = Concentration*SocialInteraction
doc  Contacts = Contacts of infected with susceptible.
aux  Error = RealData-Infected
doc  Error = Difference between real and simulated data.
aux  S_and_I = Susceptible+Infected
doc  S_and_I = Number of susceptible and infected.
aux  SocialInteraction = SocialFactor*Infected
doc  SocialInteraction = Intensity of social interaction.
const  InfectionSpreadFactor = 0.095
doc  InfectionSpreadFactor = Ratio of infected when in contact.
const  InitNumber = 1289
doc  InitNumber = Initial number of susceptible.
const  RealData = 0.51
doc  RealData = Real data of infected each day from statistics.
const  SocialFactor = 1.885
doc  SocialFactor = Average number of contact per person per day.
spec  start = 0.00000
spec  stop = 74.00000
spec  dt = 0.12500
spec  method = RK4 (fixed step)

Figure 2. Bass diffusion model adapted for epidemics spread.

3. Results
Table 1 shows the best combinations of parameters after the optimization with genetic algorithms. Genetic algorithms with 50 generations and 25 genes per generation were used. Free variables were: InitNumber of Susceptible, SocialFactor and InfectionProbability.

| Country  | Susceptible(0) | Social Factor | Infection prob. | Sf x Ip | SSE          | Days |
|----------|----------------|--------------|-----------------|---------|--------------|------|
| Austria  | 15606          | 1.735        | 0.159           | 0.275865| 1.169E8      | 67   |
| France   | 131309         | 3.674        | 0.062           | 0.227788| 8.747E7      | 91   |
| Italy    | 208881         | 0.512        | 0.521           | 0.266752| 6.176E10     | 109  |
| Korea S  | 10801          | 2.273        | 0.110           | 0.25003 | 4.275E6      | 83   |
| Slovenia | 1289           | 1.885        | 0.095           | 0.179075| 259844       | 74   |
| Swiss    | 28004          | 0.618        | 0.489           | 0.302202| NA           | 86   |
The criteria function was Sum of Squared Errors which is shown in the column SSE. The days of the first Covid-19 wave are also listed. Here the time series that resembles S-curve were taken out of the available data.

The average product of Social Factor x Infection probability is 0.250 with standard deviation of 0.043. This product characterizes the intensity of the disease spread. At start of the epidemics, that would mean, that four infected would transfer the disease to one healthy susceptible person. It is interesting, that this factor is similar for different countries which have significantly larger number of susceptible.

Figure 3 shows the results of the simulation model which was run with best parameters after optimization with genetic algorithms. On the x axis the time in days is shown and on the y axis the number of infected is represented. Red line (-1-) represents the real data and green one (-2-) represents the results of the simulation model. One can observe different values on y axis since the countries considered in the first wave of the Corona pandemic have different number of susceptible people which corresponds to the population size. This value varies from 2775 for Slovenia and 234998 for Italy. The S shaped growth can be observed in all cases, the series were intentionally selected in order to make the comparison with the Bass diffusion model.
The validation was performed on \( n \) points according to the following equations [10, 11] where the coefficient of determination is defined as:

\[
 r^2 = \left( \frac{1}{n} \sum_{t=1}^{n} \left( S_t A_t - \overline{S} \overline{A} \right) / S_A \right)^2
\]

Mean Absolute Per cent Error (MAPE):

\[
 MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| S_t - A_t \right| A_t
\]

Bias component of MSE (Mean Squared Error):

\[
 U^m = (\overline{S} - \overline{A})^2 / MSE
\]

Variation component of MSE:

\[
 U^s = (S - A_A)^2 / MSE
\]

Covariation component of MSE:

\[
 U^c = 2 (1 - r) S_A / MSE
\]

where \( A_t \) are actual numbers gained by the publically available data [1, 2] and \( S_t \) are simulation results provided by the developed model. Table 1 shows the validation of simulation runs for all countries.

Coefficient of determination is high, on average 0.97 which indicates, that with the simple Bass diffusion model, the initial estimates of pandemic dynamics could be made. However, mean absolute per cent error for all time-series is 30% so one should also consider this limitation.

Table 2. Validation of simulation runs.

| Number of data points (n) | Austria | France | Italy | Korea | Slovenia | Swiss |
|--------------------------|---------|--------|-------|-------|----------|-------|
| Coefficient of Determination (R^2): | 68 | 92 | 109 | 84 | 75 | 87 |
| Mean Absolute Per cent Error (MAPE): | 0.965 | 0.974 | 0.940 | 0.973 | 0.987 | 0.972 |
| Mean Square Error (MSE): | 206555 | 10294391 | 899142 | 452726 | 47563 | 10 |
| Root Mean Square Error (RMSE): | 1437 | 10146 | 29986 | 673 | 62 | 2181 |
| Bias component of MSE (Um): | 0.075 | 0.148 | 0.003 | 0.005 | 0.073 | 0.075 |
| Variation component of MSE (Us): | 0.174 | 0.291 | 0.352 | 0.027 | 0.289 | 0.015 |
| Covariation component of MSE (Uc): | 0.751 | 0.828 | 0.645 | 0.969 | 0.637 | 0.911 |

Figure 4 shows scatter diagram of parameter variation with genetic algorithms for six countries. The graphs are marked with letters A-Austria etc. One can observe, that in several cases the pattern in search space emerges. This might indicate the Pareto front and could further be examined.
4. Discussion
COVID-19 pandemic first wave in many case exemplifies S-shaped growth that could be roughly approximated with Bass diffusion model. In several countries the time-series follow this shape while in others there are some deviations. However, the overall dynamics is in all cases defined by S curve [4]. The model was successfully developed by the principles of system dynamics and realized with Powersim. By stating the criteria function the model parameters were determined. General spread factor for all considered countries was 0.25. Although the countries are different in size and though the number of susceptible is significantly different the dynamics is matching.

5. Conclusion
The performed validation has proved that the model could be used for rough estimates with good correlation; however, there are deviations between the model and the real system response especially in the second part of the process where the next wave is emerging.

Important aspect of model development is connection to the live database in order to be operational for public policy design, organization of medical processes [5] and effectively coping with the pandemic. Here the planning showed to be of crucial importance.
Acknowledgements
This work was supported by the Slovenian Research Agency (ARRS) within SI-MNE bilateral project “Development of a cyber-physical system for stress monitoring in endangered individuals and groups” Proj. No.: BI-ME/18-20-009, SI-RF bilateral project “Numerical-analytical examination of biologically inspired algorithms for control of Cyber-physical Systems” Proj. No.: BI-RU/19-20-034, and Research program “Decision support systems in electronic commerce”, program No.: UNI-MB-0586-P5-0018.

References
[1] South Korea: Coronavirus Pandemic Country Profile - Our World in Data 2020 https://ourworldindata.org/coronavirus/country/south-korea?country=--KOR
[2] GOV.SI Coronavirus Disease COVID-19 2020 https://www.gov.si/en/topics/coronavirus-disease-covid-19/ (Accessed 16.11.2020)
[3] Wang W, Fergola P, Lombardo S and Mulone G 2006 Mathematical models of innovation diffusion with stage structure Applied Mathematical Modelling 30 129-46
[4] Chang S, Pierson E, Koh P, Gerardin J, Redbird B, Grusky D and Leskovec J 2020 Mobility network models of COVID-19 explain inequities and inform reopening Nature Epub ahead of print 33171481
[5] Stojanovic R, Skraba A and Lutovac B 2020 A Headset Like Wearable Device to Track COVID-19 Symptoms 9th Mediterranean Conference on Embedded Computing (MECO) ahead-of-print
[6] Gurumurthy K and Mukherjee A 2020 The Bass Model: a parsimonious and accurate approach to forecasting mortality caused by COVID-19 International Journal of Pharmaceutical and Healthcare Marketing ahead-of-print
[7] Decock K, Debackere K, Vandamme A and Van Looy B 2020 Scenario-driven forecasting: modeling peaks and paths Insights from the COVID-19 pandemic in Belgium Scientometrics 124 2703-15
[8] Ku C, Ng T and Lin H 2020 Epidemiological Benchmarks of the COVID-19 Outbreak Control in China After Wuhan's Lockdown: A Modelling Study with an Empirical Approach SSRN Electronic Journal
[9] Adam D 2020 A guide to R - the pandemic’s misunderstood metric Nature 583 346-8
[10] Oliva R 1995 A Vensim ® Module to Calculate Summary Statistics for Historical Fit D-4584 System Dynamics Group (Massachusetts Institute of Technology, Cambridge MA) http://metasd.com/wp-content/uploads/2010/03/D4584theil.pdf
[11] Sterman, John D 1984 Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models Dynamica 10(2) 51-66