Continuous Calibration of an Agent-based Simulation Model for COVID-19

Bertan Imre [1] and Wish Malayandipattinamkaruppana Natarajan [2]

[1] Trier University, Trier 54296, Germany
[2] Trier University, Trier 54296, Germany

1 Abstract

In this paper we present about continuous calibration for a set of data from Robert Koch institute about the spread of contagion Covid-19 which is then run on the agent-based model created in netlogo that simulates the spread of the contagion. The model shows typical scenarios like spread of contagion before the lockdowns and after the lockdowns, in which we show calibration and its consistency with the data and how close the model is to replicate situations represented in real world. The model here is treated as a black box from which we get a set of parameters after calibration is done and new parameters for data that is being generated during lockdowns, there by generating new parameters for each calibration being done there by leading to continuous calibration, which also addresses when the calibration and recalibration with new parameters will be required.

2 Motivation

Agent-based Model (ABM), that is bottom-up simulation model, increases its popularity in different fields such as from archaeology, biology and ecology, supply chains, consumer market analysis, military planning, and economics. Moreover, various kind of disciplines, like artificial intelligence, complexity science, game theory etc., consider Agent-based Simulation (ABS) and utilize its models [1]. For all kind of simulation models including Agent-based Models (ABMs), the most significant factor is validity of the model. Correct results can be reachable as long as the model is valid and for this, calibration of the model is one of the essential steps [2].

Calibration step can be considered as optimization problem for finding most accurate parameters of the model. In this step, predetermined input and output data are used and after specifying parameters, the model can be used for forecasting, identifying insight of the system etc. However, there are several difficulties of determining parameters because of model characteristics such as non-linearity, high dimensionality as well as low quantity and quality of observed data (Oliveros-Ramos and Shin 2016). Black box and white box that deal with these problems are two main parts of calibration and various techniques of these are available for modeler according to their model [3].

In the paper we have considered our model as a black box as it best describes the situation of corona in Germany and its cities as it considers many unknowns which is exactly what is actually happening in the current world.

In this paper, we present a continuous calibration that simulate the situation in Germany for Covid-19 by using Social Simulation for Analysis of Infectious Disease Control (SoSAD). Continuous calibration is suitable for our case because the process is dynamic, and our knowledge of Covid-19 is still changing. We are taking data that is being reported on day-to-day basis and feeding it to the model based and constantly calibrating it. Now by doing continuous calibration we can get a better representation and understanding of the spread and places being affected by the Covid-19 virus from which we can device out some prevention methods in areas that could be prone to the already affected area.
Further here the need for doing calibration again is being done during the lockdowns that is because it is a situation which definitely has different set of parameters as many lockdown conditions were set like limiting daily contacts, cancellation of many public celebration events in huge numbers, closing of schools and other public affairs places. Since for our study we are using the situation in Germany for the basis of our model we are constantly supplying our model with data provided on Covid-19 sites for numbers of infected here we would face a slight challenge as numbers for the weekend from Friday after the reporting is already done till Monday's end of reporting. As for this range of days we will always get a result on Monday after the reporting is done Calibrar package in R is used because its main purpose is to fit complex models to data in our case, we are fitting an external Netlogo model in R.

3 Calibration and Calibration Techniques

Calibration is traditionally conceptualized as a step-in model validation. It involves systematic adjustment of model parameters so that model outputs can accurately reflect the actual system behavior. To calibrate a model, three important issues need to be addressed. The first issue is to select significant metrics to represent the emergent behavior of the target system and to specify a general and effective fitness function to measure the distance between a simulated scenario and the real situation. The second issue is to reduce the computation time because exhaustive search in parameter space is expensive (exponential growth with the number of parameters). The third issue is to obtain robust solutions for avoiding the over-fitting problem [7].

Calibration helps the model to be able to reproduce patterns seen in data. It is important to calibrate a model since calibration is a vital part of the modeling and simulation process and denotes the determination of parameter values by estimating them from comparison between simulation results with reference data. Standard society calibration techniques treat a society simulation model as a black box, which computes a function that cannot be written down explicitly. In general black box calibration methods try to obtain and use an approximate relationship between input and output variables of the simulation for determining the ”optimal” input setting. Some popular black box calibration approaches are gradient based search methods, stochastic approximation methods, sample path optimization, response surface optimization and heuristic search methods. An advantage of the black box approach is that it is not important for the calibration procedure what kind of simulation has to be calibrated. This advantage is also a big drawback. Since no knowledge about the internal structure and the parameter dependencies of the simulation model is used. In white box simulation calibration, we explicitly use model knowledge to enhance the calibration process [3]. Calibration process can be formulated as a series of local minimum searching problems. There are many ready-made methods for searching local minimum value of a given fitness function. For example, Powell’s method for finding a local minimum of a function [7].

Some methods for Calibration of Agent-based Simulation methods are as follows:

1) SPSA (Simultaneous Perturbation Stochastic Approximation) is chosen to be the calibration algorithm to solve the calibration problem. To minimize the difference of mode share distribution of census data and simulation output, eight parameters of the utility function in MATSim are calibrated [9].

2) Bayesian analysis system of statistical inference based on interpreting probability to model calibration, specifically, approximate Bayesian computation. While experiments have previously been done on this approach within the context of agent-based models [8].

3.1 Re-calibration and Continuous Calibration

Calibration in simple terms is to get a nearly accurate range for a given task by observing the data at hand and in this paper we further discuss with the daily data that’s constantly updating and hence we need to calibrate our model to fit the actual occurrence that’s happening by adjusting the factors given to us, but the new data always changes and brings in new numbers due to this we face the challenge of re-calibration and continuous calibration because it is an on-going process. These changes are shown in the simulation model. Re-calibration simply put can be said as “to perform calibration again with new parameters so that the result will always remain to nearly accurate”. Similarly, we can then define Continuous calibration in simple terms as “the process of calibrating and re-calibrating a number of times so that the result always remains near to accurate.” [5]. With this in place we proceed with further details about the Social Simulation for Analysis of Infectious Disease Control (SoSAD) model which is in place as our case study.

In our model we have the data of number of covid-19 cases which we get daily and various other factors. Now as we already have quite a lot of data that is already available, we can use that as observation data or test data. This data is generated on daily basis which requires re-calibration of our model and continuous calibration of our model since we cannot stay focused on same test data every time, hence we can say that our test data is of evolving nature as its generated daily depending on the number of covid19 cases per day. We focus on weekly data as we will have the challenge for weekend logic which is absence or reporting on holidays. For now, which we are considering it to be a moving average or determined time interval data which is discussed later. So for this we re-calibrate our model every day till Friday of current week so that we get a near to accurate result which will help us for the weekend calculation of a moving average or determined time interval data, then we replace it with actual factual numbers on Monday. The logic can determine for data which seems to be missing. Different set of parameters are already present in our model such as number of touches, type of people, behavior of people, parameters of medical infrastructure, disease progression etc. all of them are vital for the knowledge of spread about the contagion.

With this in place we get to know the importance of re-calibration of model because, we will at a point have the calibration properties of the model for a specific time then with new data incoming every day then the same properties will change and may develop new ones. Due to this we need to re-calibrate our model continuously which will indeed help us at the end for the mainly taking decisions factors like “Should schools or universities be closed?”, “reduction in social interaction”, “How are children and adults’ free time managed?”, “A point where its complete lockdown – No schools open, high level of caution followed by the people in affected areas, only essential businesses open etc. During this again we will have our parameters changed which will again affect our model output” [6] so from this we get a timeline of events which will fit to the parameters of what to be put in the model to get a better understanding. This is described in detail further in the paper.

3.2 Continuous calibration concept.

As mentioned before calibration is a process of running a model using a set of input data and then comparing the results of the experiment with the actual measurements that are obtained from the system and this process is repeated for a given model with various parameters till satisfactory results are available. Now this leads us to a gray area where we keep achieving satisfactory results but till what time is the actual question? And it is certain that external parameters can change any time after some careful thought it was necessary to do calibration continuously with different sets of parameters to achieve satisfactory/valid results.

To simply put our concept fig1.1 shows the flow of the continuous calibration.

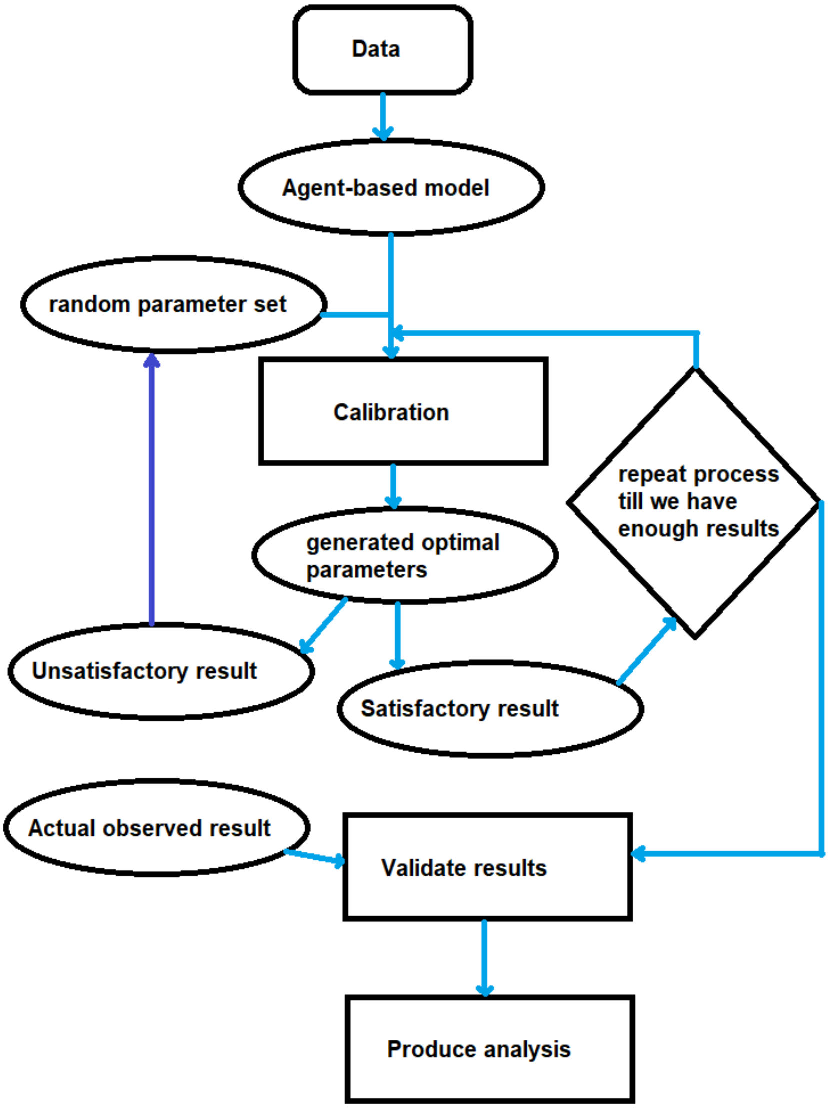

Fig1.1

Here it is considered as random parameter set as we treat the model as a black box. With this the concept of why continuous calibration is important becomes clearer. The continuous calibration will ensure that our system will be optimally calibrated and produce results efficiently and will re-calibrate again if unsatisfactory results appear. We determine our satisfactory results by using the MSE (mean squared error) and CI (confidence interval)/forecast range. First of all, the command function from which the data is fetched to the agent-based model, on which some random parameters have been chosen so that the calibration simulation can be performed on the model. Once this part completes and we have enough satisfactory results we generate the MSE from the actual data and the results after enough number of runs this will give us our accuracy of the model how good the results have been generated from the simulation runs. Post this we calculate a confidence interval/ forecast range to how low and how high the new test values can be obtained from the model for future analysis and validation of the simulated calibration that is being performed. The agent-based model and operations that have been carried out on it with example of SoSAD (covid-19) are discussed further in this paper.

3.2.1 Blackbox optimization for continuous calibration.

The black box optimization techniques are effective when it comes to unknown parameters which is exactly the case in our model because of this it is quite effective in generating our satisfactory results as depicted in fig1.1.[13]

Hence here we are concerned with the problem of determining what and all effects of different parameters on the output of a model how the vary or do they remain similar. This problem lies at the heart of all science for any optimization of the process [12]. Let us us consider a simple example “the human body” treating this as a black box of study can tell us various responses to various situations, they can be same or different, but till what time will the result be similar or when will it change is something that cannot be effectively suggested without much data about the parameters of conditioning the experiment. This was just a simple explanation about how challenging the concept of black-box approach can be but is highly effective as we introduce the parameters step by step to achieve to a conclusive satisfactory result.[12]

So, in general the black box calibration is good to derive approximate results between input and output variable that are in our system. This talks about the advantage of black box technique that it is not important for the calibration procedure to know what type of simulation is being calibrated, but as this may be its advantage can also be its disadvantage, as not much information about the simulation model is know from the beginning there by making the algorithm run quite longer to get significant results for data analysis and validation of results.[3]

4 Netlogo and R for calibration simulation.

Our project deals with an external Netlogo software for the simulation model and we use R for the analysis part that is parameter estimation, and calculation of MSE and forecast range/CI. The ability to generate accurate parameter estimation has been used as the criteria to check the usefulness of ecological models (Bartell 2003). Consider any model, then the criteria to check for the best possible parameter set would be the optimization of the scalar objective function with respect to the model parameters (Walter and Pronzato 1997, Bolker et al. 2013). Hence after the objective function has been created then the parameter estimation is just an optimization problem waiting to be solved. In our case this is going to be a difficult task as the model characteristics like non-linearity, high dimensionality, treating the model as a black box, and less prior information on already existing parameters fall at hand (Tashkova et al. 2012). In addition, our model is a complex ecological model for SoSAD which is numerically intensive and requiring long simulation runs to complete the simulation add up to the difficulty [13].

Here we use the r package called calibrar that is specifically designed for parameter estimation of complex ecological models. It also fulfils our need of treating the model as a "black-box" approach, this packages advantage is that, that it allows to use models implemented in any language with r. It also provides us with a generic interface with models that allow us to help in the construction of the objective function in r, without doing any changes in the model itself [13].

Function	**Returned Objects**	**Description**
calibrate	An object that is summarizing the calibration results	Performs a sequential parameter estimation of a model using multiple phases
createObjectiveFunction	A function, integrating the simulation of the model and the comparison with observed data	Create a new function, to be used as the objective function in the calibration, given a function to run the model within R, observed data and information about the comparison with data

Table 1.1[13]

The general workflow of the calibrar r package is shown in fig1.2.

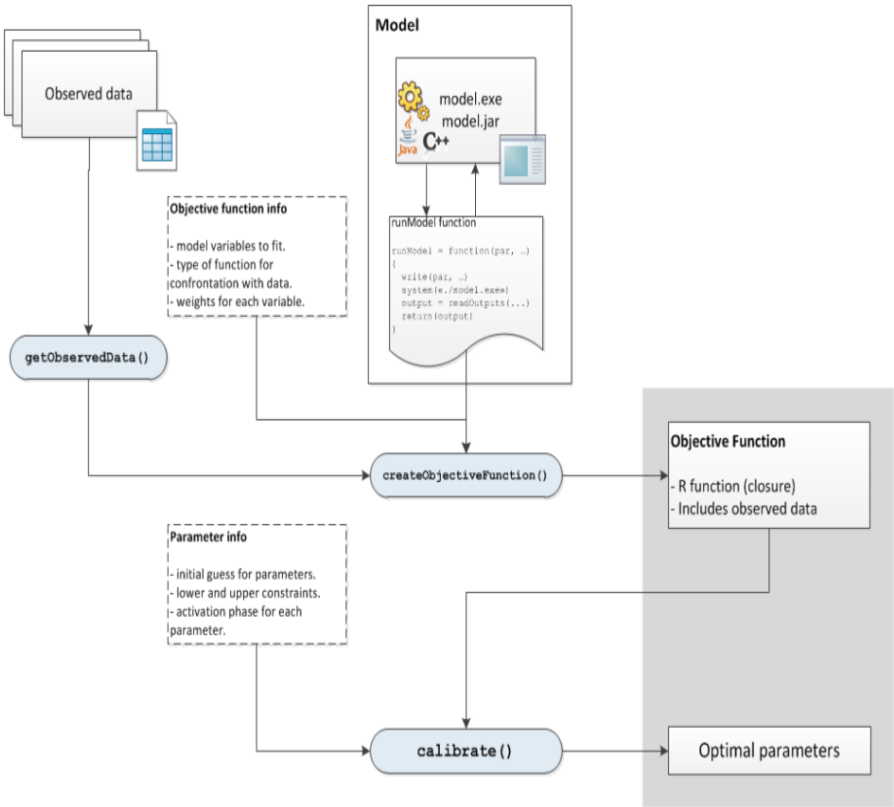

Fig1.2[13]

General focus is to create an objective function that can help minimize the model and thereby use it in the next step. After the objective function is created, we need to use the calibrate method to generate the optimal parameters. General use of the calibrar methods createObectiveFucntion() and calibrate() are explained in table 1.1[13].
After we have successfully generated the optimal parameters, we proceed to the validation part of results from which we generate our confidence interval/ forecast range that will help us in the test data part. Further in detail description about the calibrar r package can be found in the references.[13]

5 Application.

In this part we consider Covid-19 pandemic situation in Heinsberg, Germany. An existing agent-based simulation model, SoSAD - Social Simulation for Analysis of Infectious Disease Control, was used to simulate the situation [14].. Behavioral parameters of the model, that have significant effect on the spread of the virus, are the calibrated parameters of the model via black box calibration technique. We showed how continues calibration is applied with the new data.

5.1 Covid-19 in Germany.

First official case was confirmed at the end of January 2020 in Germany. Even if the government decided some restrictions, like closing shops, restaurants, bars etc., switching online education, closing borders, the number of cases increased. [15].
Until now (March 2021), there has been 2 waves of the pandemic covid-19. The first one started with the beginning of the pandemic and ended beginning of May 2020. The second one started beginning of October 2020 and still continues. Between the waves, the spread of the virus was relatively low and stable [16].
During the pandemic, there were “hotspots” specifically beginning of the pandemic. In these areas, the numbers of the covid-19 cases dramatically increased. After, the virus spread throughout Germany and these so called “hotspots” have been changing during the entire time frame.
One such hotspot was the city of Heinsberg. We choose this particular “hotspot” and use the data from the beginning of March 2020 to the end of June 2020 (4 months). The reason why we choose this hotspot is because the model is more appropriate for a city. Also, this hotspot is suitable to show increasing trends in a more capable manner. Moreover, the model is capable to apply lockdowns once.
Fig 2.1 shows us the percentage of cases according to Heinsberg’s population from 01.03.2020 till 30.11.2020.

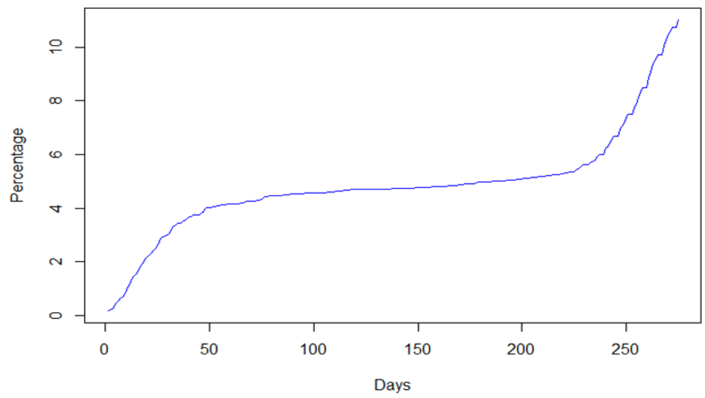

Fig 2.1

5.2 Covid-19 Model.

In SoSAD - Social Simulation for Analysis of Infectious Disease was used as an agent-based simulation model. The stochastic model was coded in NetLogo. There are several parameters that affect characteristic and spreading speed of the pandemic.

In our work, we focused on the number of cases during time. Therefore, we centered the behavioral parameters of the model, that includes work contacts, school contacts, free time contacts for children, free tie contacts for adults and maximum contact during restrictions. During the calibration steps, we tried to find the best parameter values that gives the closest results to real data.

Another important parameter group for the number of cases is distancing measures, that includes restriction and reallowing contacts, closing and reopening schools with reopening rate, home office starting and ending dates with home office rate. For these parameters, we followed the regulation of German government. Moreover, we assumed that ill people stay at home.

We kept default other parameters groups like demographic parameters, parameters of medical infrastructure, infectiousness and outbreak, disease progression. In addition, vaccination didn't play a role in our model because of the time we considered. In our simulation model the population size was 1000 and the real population is 42236.

As mentioned previously we have ran the simulation model on R programming language via rJava and RNetLogo packages to use calibrar package in our calibration process.[13].

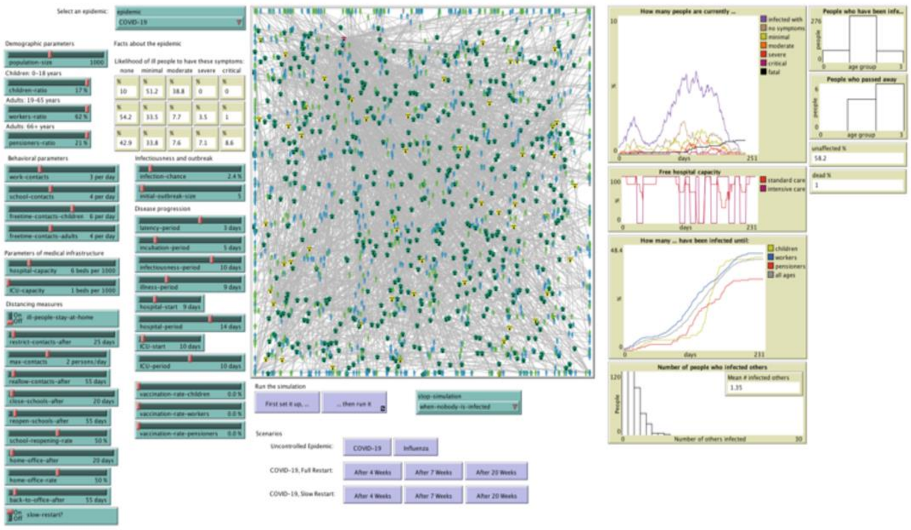

Fig 2.2

SoSAD - Social Simulation for Analysis of Infectious Disease interface

5.3 Model calibration.

We wrote a function to run and calibrate the simulation model. The function takes 4 parameters that are parameter set, parameter names, number of repetition and real data.

covid_simulation_train <- function (param.set, param.names, no.repeated.sim, real.data)

Parameter names and their values, that are calibrated, are represented by "param.set" and "param.names". The function should be run several times because of its stochastic characteristic and the user can chose the number of repetitions via "no.repeated.sim". Lastly, the user can use real data via "real.data".

The function takes parameter values and set the behavioral parameters of the simulation function. Likewise, the fixed values for the distancing measures are assigned. The function runs several times and calculates average percentage of the cases day by day within predefined range. Finally, the function compares the real values and the values that the simulation model gives and returns mean squared error (MSE).

Our purpose was minimizing the error and tried to find the closest results to real data. For this goal, we used three methods of calibrar package, "cmaes", "ahr" and "L-BFGS-B". These three methods are appropriate for our model because they allow to restrict the parameters with lower and upper bounds. We used these bounds to escape meaningless results, like minus contact.

We used first 21 days as training data and tried to find best parameters that fits the real situation. According to the MSE values of these three methods, we decided to use “cmaes” method to forecast the future values and calibrate the parameters according to validation data.

	MSE	Work-contacts	School-contacts	Freetime-contacts-childern	Freetime-contacts-adults	Max-contacts
cmaes	0,0185	3,23	3,95	3,93	3,36	3,37
lbfgsb	0,1738	3,30	3,30	3,02	3,30	2,91
Ahr	0,5999	5,16	5,71	2,14	2,13	3,41

Table 2.1

Table 2.1 shows the MSE values and the values of the parameters as calibration results of “cmaes”, “ahr” and “L-BFGS-B” methods.

5.4 Test of the model and decision.

After the first calibration step, the parameters can be used to forecast the future situations. The data that the simulation function creates can give insights about future trends. Moreover, the parameters should be validated and calibrated continuously because of catching new trends in the pandemic.
For this reason, after we got the real data, we used the function again and created 95% of confidence interval for approximately 60 days after the first train days for validation process. If the real data exceed lower or upper bounds, it means that the simulation function is not good enough to forecast future trends and it should be recalibrated.
As it is seen in the figures, prediction of the first parameters work for first 29 days but after the real data exceed upper bound. Therefore, we ran the simulation function again up to that date and found new parameters. This step is second calibration step of the continuous calibration (recalibration). After we got the real data again, we did second validation and calibration step. This time, the simulation function worked well for 45 days more after first exceeding the confidence interval.
After, decreasing trend was observed and it means that the simulation function should be recalibrated again to catch the decreasing trend. During the pandemic this continuous calibration process can be used to catch new trends.

Table 2.2 shows the MSE values and the values of the parameters as calibration results of the first and second calibration.

	MSE	Work-contacts	School-contacts	Freetime-contacts-childern	Freetime-contacts-adults	Max-contacts
1st	0,0185	3,23	3,95	3,93	3,36	3,37
2nd	0,0137	4,22	5,70	2,95	3,16	1,58

Table 2.2

According to second calibration (recalibration), school contacts dramatically increased and maximum contact (under lockdown) significantly decreased.

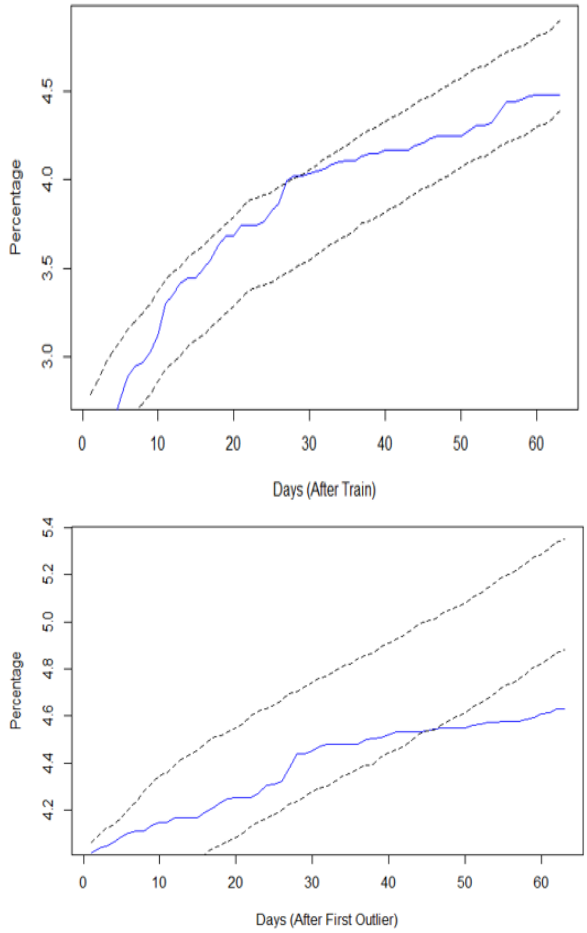

Fig2.3

The first graph shows the confidence interval of the simulation function and the real data after train days (starts from 22.03.2020) and the second graph shows the confidence interval of the simulation function and the real data after first outlier (starts from 20.04.2020).

6 Conclusion.

Agent-based modeling is a way to understand the components of a real situation and their relationship with each other via bottom-up simulation. Calibration of the parameters of a simulation model is one of the issues in the agent-based modeling process. Moreover, in some unstable cases like pandemics, the parameters should be calibrated continuously to adapt the model to real situations. Black-box techniques might be useful when a user of the techniques does not have enough knowledge to interpret the situation that he or she is interested in or the situation is totally new. If black-box techniques are used in a continuous manner, the user could observe changes of the parameters and reach insights about the situation in a much better way.

References

1. Kasaie & Kelton: Guidelines for design and analysis in agent-based simulation studies. In 2015 Winter, Simulation Conference (WSC), pp. 183-193, IEEE (2015).
2. Klügl: A validation methodology for agent-based simulations. In Proceedings of the 2008 ACM symposium on Applied computing, pp. 39-43 (2008).
3. Fehler, Klügel & Puppe: Techniques for Analysis and Calibration of Multi-Agent Simulations. In Engineering Societies in the Agents World V, 305-21 (2005).
4. Singh, K., & Upadhyaya, S.: Outlier detection: applications and techniques. *International Journal of Computer Science Issues (IJCSI)*, *9*(1), 307 (2012)
5. Aron Fink, Sebastian Born, Christian Spoden & Andreas Frey: A continuous calibration strategy for computerized adaptive testing.
6. Miguel Casares, Hashmat Khan: A Dynamic Model of COVID-19: Contagion and Implications of Isolation Enforcement
7. Zhengchun Liu, Dolores Rexachs , Francisco Epelde , Emilio Luque : A simulation and optimization based method for calibrating agent-based emergency department models under data scarcity. Computers & Industrial Engineering Volume 103, January(2017).
8. LDRD Annual Report: Approaches for Calibrating Agent-Based Models to Data
9. Hong Liang: Calibration of Agent-based Transport Simulation with SPSA method. Spatial Development and Infrastructure Systems. July (2018).
10. Continuos calibration optitrack: https://v22.wiki.optitrack.com/index.php?title=Continuous_Calibration

11. Forrester AI. Black-box calibration for complex-system simulation. Philos Trans A Math Phys Eng Sci. 2010 Aug 13;368(1924):3567-79. doi: 10.1098/rsta.2010.0051. PMID: 20603368.
12. Forrester, Alexander. (2010). Black-box calibration for complex-system simulation. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences. 368. 3567-79. 10.1098/rsta.2010.0051.
13. Oliveros-Ramos, Ricardo & Shin, Yunne-Jai. (2016). Calibrar: an R package for fitting complex ecological models.
14. SoSAD- Social Simulation for Analysis of Infectious Disease Control: Uni Trier: https://www.uni-trier.de/index.php?id=73127
15. COVID-19 Pandemic in Germany Wikipedia: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Germany#Government_reactions
16. Robert Koch Institute Covid-19 Dashboard: https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4/page/page_1/