



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Lecture with Computer Exercises:
Modelling and Simulating Social Systems with MATLAB

Project Report

**Civil Violence: Modelling Genocides
Model Extension and Global Sensitivity Analysis**

Ruben Wälchli & Zoran Bjelobrk

Zurich
December 10, 2014

Agreement for free-download

We hereby agree to make our source code for this project freely available for download from the web pages of the SOMS chair. Furthermore, we assure that all source code is written by ourselves and is not violating any copyright restrictions.

Ruben Wälchli

Zoran Bjelobrk

Contents

1 Abstract

2 Individual contributions

3 Introduction and Motivations

Ethnical conflicts were omnipresent throughout human history and still are a major source for civil violence in these days. The wars leading to the decay of former Yugoslavia are an example in the recent past. When studying the project suggestions, this topic immediately drew our interest.

Our project is a so-called "model-driven" project. The base model was described by Epstein in 2002. It contains some very strong simplifications as further described in section ???. Our approach was to modify the model in a way that describes reality better in our opinion and to compare the results of the modified model runs to those of the base model presented in literature. To further assess the models in a quantitative manner, it was tried to perform a global sensitivity analysis for most of the model parameters based on a method described in literature.

The main questions of this work were: Is it possible to relax some of the assumptions made in literature and introduce new interactions to make the model more realistic? Can the effects of the model parameters on the outcome of the simulation be quantified, both for the base and the modified model? Having identified the most influential parameters, what is the dependence of the simulation outcomes on them? Can in the end, based on those results, a statement be made about the qualitative and quantitative difference between the base and the modified model? Does the modified model actually describe the reality better than the base model?

4 Description of the Model

4.1 Base Model

The model is a so-called agent-based model. There are two fundamental classes of agents: civilians and law enforcement officers (LEOs). The only purpose of the LEOs is to prevent violence through arresting active, this means violent, civilians. The civilians are further divided into two ethnical groups. They become active or stay quiet (peaceful) based on the values of their individual characteristics and the environment within their range of sight. The agents are placed on a two-dimensional map on which they can move and interact with each other.

4.1.1 Specification of the Civilians

The will of a civilian to become violent is described by its grievance G . This individual grievance is described by two components: the hardship H of the civilian and his/her perceived legitimacy of the other ethnic group L .

$$G = H(1 - L) \quad (1)$$

The grievance of the individual civilian is according to this formula given by the hardship multiplied by the "illegitimacy" $(1 - L)$ of the other ethnic group. The hardship is heterogeneous across the civilians. It is initialized for each civilian at the beginning of the simulation by drawing it from the uniform distribution on the interval $[0, 1]$ ($U(0, 1)$) and it remains constant over the course of the simulation. L is homogeneous across all civilians and also stays constant. It is also equal for both ethnic groups. It is defined at the beginning of the simulation and also lies between 0 and 1.

R is the risk aversion of the civilian. It is heterogeneous across the population and also drawn upon initialization from $U(0, 1)$. A civilian inspects its environment before deciding whether to go active or stay quiet (and vice versa). The civilians have the vision v , which is homogeneous across the population and is defined at the beginning of the simulation.

$$P = 1 - \exp \left(-k_P \left(\frac{LEO}{A} \right)_v \right) \quad (2)$$

P is the arrest probability estimated by the civilian. $(LEO/A)_v$ is the LEO-to-active ratio within the civilians vision. A is always at least 1, because the civilian counts himself as active when calculating the arrest probability. The reasoning behind this calculation is, that it is less likely for the civilian to be arrested, when already a lot of other civilians are acting violently and very few LEOs are present around him. The opposite can be said, if there are a lot of LEOs and only few or no active civilians. k is a constant parameter, that is calculated by setting $P = 0.9$ when $(LEO/A)_v = 1$. With the risk aversion and the estimated arrest probability the so-called net risk of the civilian of going active can be calculated:

$$N = RP \quad (3)$$

The difference $G - N$ is the civilian's net utility of becoming active. This is compared to a threshold value T . If $(G - N) > T$ the civilian becomes or stays active and if $(G - N) \leq T$ the civilian stays or becomes quiet. The threshold value T is equal for all civilians and defined at the beginning of the simulation.

4.1.2 Specification of the LEOs

The LEOs have only one characteristic, their vision v^* . They inspect the map within their vision and randomly arrest an active civilian.

4.1.3 Evolution of the Simulation

After choosing values for the global parameters, the three different types of agents are placed on the map randomly and their individual attributes are initialized. The number of civilians and LEOs that are placed on the map are determined by their corresponding selected densities. In each iteration begins by the random selection of one agent (civilian or LEO). This agent moves to a random empty position within his vision. There he inspects his environment. If the agent is a civilian he takes the decision to be active or quiet. When he decides to become active, he tries to kill a random civilian of the other ethnic group within his vision. If the selected agent is a LEO, he arrests a random active civilian within his vision. The arrested civilian is assigned a jail term and moved to the jail. The jail term is drawn from $U(0, J_{\max})$. The ages both of the civilians on the map as well as of those in jail are updated and the civilians that have reached their maximum age are removed. The number of turns of the inmates spent in jail are updated and those who have served their time are released onto random positions on the map.

4.2 Modification of the Model

4.2.1 Individual Perceived Legitimacy and Threshold

In our opinion, the assumption that the perceived legitimacy of the other ethnic group and the threshold value are identical for all individuals of the population is too harsh. Therefore, we wanted to make them individual quantities. To us it also seemed not realistic to have the perceived legitimacy uniformly distributed throughout the population. Therefore, we decided to use the Gaussian distribution truncated to the interval $[0, 1]$. In the modified model the perceived legitimacy of the other ethnic group is initialized for all civilians by drawing it from this distribution. This adds two parameters to the model: the mean μ_L and the standard deviation σ_L of the legitimacy.

The threshold T can be seen as the level of "extremism" an individual has. Epstein describes it as the utility of not publicly expressing one's private grievance. A very extreme individual will see absolutely no utility in holding his private grievance back, whether a very conservative individual will almost never publicly express his

private grievance. The difference ($G - N$) on the other hand is the perceived utility to express one's private grievance publicly and it can take values in the interval $[-1, 1]$. A civilian with $T = -1$ will always be active, whereas a civilian with $T = 1$ will never be active. As said before, it is probably unrealistic to have the threshold value uniformly distributed over the population. Therefore, it is also drawn upon initialization of the civilians from the truncated Gaussian distribution on the interval $[-1, 1]$ adding another two parameters to the model (μ_T and σ_T).

For the sake of simplicity and to make the sensitivity analysis described later not computationally too expensive, it is renounced to also change the uniform distributions of the other individual characteristics of the civilians (H and R) to truncated Gaussian distributions.

4.2.2 Updating the Perceived Legitimacy

It seems quite obvious that the perceived legitimacy of the other ethnic group should not be a constant over the course of the simulation. Violence committed by one ethnic group will probably decrease its perceived legitimacy in the eyes of the other ethnic group. The opposite effect will possibly have the arrest of a active (violent) civilian on the perceived legitimacy of its ethnic group. In our expansion of the model, these effects are limited to the range of sight of the involved killed or arrested active civilian. The update in the case of violence is the following:

$$L_{\text{new}} = L_{\text{old}} (1 - k_L) \quad (4)$$

When an active civilian is arrested, the update is given by:

$$L_{\text{new}} = L_{\text{old}} + k_L (1 - L_{\text{old}}) \quad (5)$$

The reasoning behind both formulas is that if the perceived legitimacy is either very high or very low, violence and arrests only change little in the legitimacy (saturation effect). On the other hand, when a civilian is unsure, if the other ethnic group has the right to exist (medium value of L) such events can have a strong influence.

4.3 Global Sensitivity Analysis of Model Parameters

In order to test the sensitivity of the output of a generic model with respect to its inputs, a variance-based global sensitivity analysis can be performed as proposed in literature. For the sake of simplicity, we restrict ourselves to scalar model outputs. Most generally speaking, the model is a arbitrary relation $Y = f(X_1, X_2, \dots, X_k)$ between the output variable Y and the k input variables (parameters) X_i . Variance-based sensitivity analysis methods provide a quantitative measure how the variance

of the output depends on the different uncertain/random inputs. The main advantages of variance-based global sensitivity analysis techniques are their abilities to describe the influence of each input parameter over its full range of values and to detect interactions between the input parameters. Their main disadvantage is their computational cost.

For each input parameter X_i , a first-order sensitivity index S_i and total-effect index S_{T_i} can be defined:

$$S_i = \frac{V[E(Y|X_i)]}{V(Y)} \quad (6)$$

$$S_{T_i} = \frac{E[V(Y|\mathbf{X}_{\sim i})]}{V(Y)} = 1 - \frac{V[E(Y|\mathbf{X}_{\sim i})]}{V(Y)} \quad (7)$$

$V(Y)$ is the unconditional variance of Y , $E(Y|X_i)$ the expectation value of Y conditioned on X_i and $V[E(Y|X_i)]$ the variance of the conditional expectation (fixed/known value of X_i). This is measure for sensitivity to changes in the parameter X_i . After normalization to the unconditional variance $V(Y)$ one obtains the first-order sensitivity index, which is also called the main effect of X_i .

$V[E(Y|\mathbf{Y}_{\sim i})]$ on the other hand is the output variance $V(Y)$ conditioned on all input variables except X_i , i.e. $\mathbf{X}_{\sim i}$. This is the remaining variance of the expectation value, if the true values of all input parameters except X_i would be known.

The values both of S_i and of S_{T_i} have to be by definition between 0 and 1 as well as S_{T_i} has to be larger or equal to S_i . The bigger S_i and S_{T_i} are, the higher is the effect of the input variable X_i onto the output variable Y . The difference $(S_{T_i} - S_i)$ provides information on how much X_i interferes with the other input parameters. If it is zero, there are no interactions. For a purely additive model is $\sum_{i=1}^k S_i = 1$. The difference $1 - \sum_{i=1}^k S_i$ gives a general measure of the amount of interactions in the model. Further information about this method of global sensitivity analysis can be found in the literature.

4.4 Qualitative Assessment of the Input Parameters

Unfortunately, the global sensitivity analysis yielded no useful results. Therefore, an alternative had to be found in order to assess the influence of certain input parameters on the simulation outcomes and in the end discuss the differences between the two models. The focus was put on the changed procedures and the newly introduced interaction as well as the comparison of the two models with each other.

5 Implementation

The whole implementation is based on arrays of structs. This way of implementation may not be the fastest possible, but it is in our opinion the most intuitive one.

5.1 Base Model

The fields of the struct used to describe the agents are shown in table ?? . The central part of the program is the map, an array of size $(N_{\text{map}} \times N_{\text{map}})$ containing this struct as entries, where N is the map size. The individual values of the fields are initialized at the beginning using the function *fun_init_map()*, which returns the initialized map array with all the civilians and LEOs. Type equal to 1, 2 or 3 indicates that the agent is a civilian of the ethnic group 1, group 2 or a LEO. For a LEO only the field *v* is also different from the default value. The jail is an array of size $(J_{\text{max}} \times 1)$

Table 1: The fields of the struct used to describe the single agents and their meaning and default value.

Field Name	Quantity	Default Value
type	Type of the Agent	0
H	Agent’s Hardship	0
L	Agent’s Perceived Legitimacy	0
v	Agent’s Vision	0
T	Agent’s Violence Threshold	0
R	Agent’s Risk Aversion	0
state	Agent’s State	0
age	Agent’s Age	0
life_exp	Agent’s Life Expectancy	0
sentenced	Agent’s Jail Term	0
served	Agent’s Served Time in Jail	0

with the struct from table ?? as entries. The reasoning behind the size of the jail array is that it has to be large enough that theoretically in every iteration a civilian could be send to jail. The function *fun_movement()* moves the randomly selected agent, updates the map accordingly and returns the new position of the agent. The agents cannot penetrate the borders of the map. If the selected agent is a LEO, the function *fun_action_LEO()* inspects the fields within vision of the new position and if active civilians are present there, randomly moves one of them to the jail. To do so, it first assigns the civilian a random jail term (field ”sentenced”). Then copies the civilians struct entries to an empty position in the jail and erases the civilian

from the map by setting the struct fields at its position to the default values.

If the selected and moved agent is a civilian, the function *fun_action_civ()* is called. It first inspects the surrounding of the new position and determines the current value of P and with this the net risk N . The grievance G is also computed and it is checked if $(G - N)$ is larger or smaller than T . If it is larger, civilians of the other ethnic group within vision are searched. When there is one present, he is killed by setting all the values of the struct at his position to the default values. If there is no such civilian present, the agent remains quiet although he wants to become active. If $(G - N)$ is smaller than T , the selected civilian remains or becomes quiet in any case. When the civilian was initially active, wants to stay active but does not find a opponent to kill, he however stays active. The function returns the changed map and if there has been a kill.

After moving the agent and performing its action, the function *fun_clone()* is called. It goes through all civilians on the map in a random order. For each civilian it checks, if there are empty position right next to him. When there is an empty position, the civilian is cloned with probability P_c onto this position by copying the struct of the selected civilian. The age of the newly created civilian is reset to 0.

The *fun_update_map()* function updates the ages of the civilians on the map and removes those who have reached their life expectancy by resetting the struct on those positions to the default one.

The jail is updated using the function *fun_update_jail()*. First it updates the ages and the served times of the civilians in jail, then it removes civilians that have reached their life expectancy and releases the civilians that have served their sentence onto random empty positions on the map. The structs of those civilians in the jail array are reset to the default values. The civilians are released in the quiet state, all other characteristics remain unchanged.

5.2 Modified Model

5.2.1 Heterogeneity of the Parameters L and T

The function *truncate()* is used in order to obtain two truncated probability density object. The normal probability density objects are first created with the desired mean and standard deviation using the function *makedist()*. So in the case of the modified model, the function *fun_init_map()* takes instead of the two parameters L and T the four parameters μ_L , σ_L , μ_T and σ_T . The function *rand()* is then used to

initialize the individual values of L and T of the civilians from the truncated normal probability density objects.

5.2.2 Update of the Perceived Legitimacy

The function *fun_action_civ()* of the modified model records the position and the ethnic group of the possible killed civilian. Then it inspects the neighborhood within vision of the killed civilian and records the civilians of the same ethnic group. Their perceived legitimacy is then updated according to equation (??).

5.3 Sensitivity Analysis

For both models, the output variable describing the extent of violence that occurred during the simulation was defined as the cumulative number of kills divided by the initial population size. The normalization to the population size was done in order to investigate also the initial population density as input parameter of the models. The parameters investigated and their intervals are shown in tables and .

Table 2:

Number	Parameter	a	b
1	ρ_{tot}	0.1	0.9

5.3.1 Sample Generation

The size N of the base sample generated was 100. The necessary $(N \times 2k)$ base sample matrix of random numbers, where k is the number of parameters investigated, was generated as suggested in the literature using the *sobolset* class, which is an implementation of Sobol’s random number generator in MATLAB. The procedure to obtain the matrix from a created point set object of this class was directly taken from the documentation of the software. This matrix was divided as described in literature into the two matrices \mathbf{A} and \mathbf{B} , which in turn were used to construct the k matrices \mathbf{C}_i . The matrix \mathbf{C}_i corresponds to the generic parameter i and contains all columns of \mathbf{B} except the i^{th} , which is taken from the i^{th} column of \mathbf{A} . The sample generation was done inside the program *sample-generation* and the created sample matrices were exported into .txt files.

5.3.2 Model Output Generation

The implemented models were integrated each into a function called *fun_model()* which takes the defined input parameters as arguments and returns the normalized cumulative number of kills. This function was used to generate the output values corresponding to the sets of input parameters contained in the rows of the matrices **A**, **B** and **C_i**. Sobol's quasi-random number generator produces numbers between 0 and 1. Therefore, the input parameters had to be converted inside the functions *fun_model()* in order have their values in the desired ranges. This was done using interval mapping, which here should be illustrated shortly:

$$\begin{aligned} x \in [0, 1] &\longmapsto x' \in [a, b] \\ x' &= (b - a) \cdot x + a \end{aligned}$$

For input parameters that should take integer values, the MATLAB function *round()* was used. The interval $[a, b]$ was extended by 0.5 on both ends for those parameters in order not to decrease the probability of having the parameter values a and b . The output values were stored in the $(N \times 1)$ vectors **y_A**, **y_B** and **y_{C_i}**. All of this was done inside the program *output_generation*.

5.3.3 Sensitivity Indice Computation

5.4 Qualitative Assessment of the Input Parameters

After the global sensitivity analysis failed by yielding no sensible results, a more qualitative approach was taken to nevertheless be able to discuss the effect of certain input parameters on the simulation outcome, both of the base and the modified model. For the base model it was mainly tried to reproduce the observation made in literature regarding the dependence on the density of LEOs on the map. The values of the other model parameters that were fixed for this and the following examinations are shown in table ???. In this frame it was renounced to normalize the cumulative number of kills to the initial population size, because this parameter was never varied. The initial LEO/civilian ratio was varied from 0 to 0.1 in steps of 0.004. The model was run 10 times at each input parameter value. The mean and the standard deviation of the output variables were taken. This is not an exact reproduction of the analysis done in the reference. There the number of iterations was set to 15000 and for each value of the initial LEO density the simulation was run 50 times. As measure for the extent of violence the time until one ethnic group is completely exterminated was taken (it was set to 15000, if there was no extinction of one group). In our opinion, this is not a very good way to assess the violence during the simulation. Very violent conflicts can never lead to the extinction of one ethnic

Table 3:

Parameter	Value
Size of the Map	(40×40)
Max. Age	200
Initial Pop. Density	0.7
Initial Fraction of Ethnic Group 1	0.5
Cloning Probability	0.025
Vision of the Civilians	2
Vision of the LEOs	3
k_P	2.3
Max. Jail Term	30
Legitimacy	0.8
Violence Threshold	0.1
Number of Iterations	200

group, but to very high number of casualties. Also this method was no feasible for us, given our computational resources. Therefore, we decided to let the simulation run for one generation of civilians.

An other minor difference to the reference is the selection of the LEO density. In the reference model, the LEO density on the map can be set independently from the population density of the civilians. In our implementation of the model, only the initial LEO/civilian ratio can be selected. This was mainly done in order to prevent situations where there are too many agents to be placed on the map upon initialization. However, this leads only to minor differences, the general trends should not be affected by this.

For comparison, the same analysis was done using the modified model. In this case μ_L was set to 0.8, μ_T to 0.1 and both σ_L and σ_T to 0.1. The parameter k_L was set to 0.1. The value for the standard deviations of those parameters was selected by evaluating the broadness of the distributions by eye and selecting a sensible value. The effect of k_L will be discussed later and for the time being its value was just a guess.

6 Simulation Results and Discussion

6.1 Global Sensitivity Analysis

The computed sensitivity indices of the input parameters of the base and the modified model are presented in tables ?? and ??, respectively. It is right away obvious that those values cannot be correct. For instance, the S_i have to take by definition values in the interval $[0, 1]$ and the sum of the S_i has to be equal to 1. Also, S_i^T has to be larger or equal to S_i . To our knowledge, we have made no mistake in the implementation of the global sensitivity analysis in MATLAB. One possible explanation for the failure could be an insufficiently large base sample, leading to inaccurate estimates of the sensitivity indices. But it goes past our computational resources to increase the sample size significantly in order to check, if that is actually the case.

Table 4: Sensitivity indices computed for the base model input parameters.

Parameter	S_i	S_i^T
1	0.4748	0.2421
2	-0.0147	-0.7522
3	0.0464	-1.0529
4	0.0360	-1.1220
5	0.0810	-1.1282
6	0.0640	-1.1080
7	0.0559	-1.1037
8	0.0765	-1.0351
9	0.3150	-0.0658
10	0.1213	-0.9992

6.2 Qualitative Assessment of the Input Parameters

6.2.1 Reproduction Literature Results

The result of the analysis regarding the initial LEO/civilian ratio is shown in figure . The anticipated dependence of the mean cumulative number of kills on the LEO/civilian ratio was observed. However, the standard deviation did not increase as one might expect from the results presented in literature. This might be due to the different violence measure used or the effect of the initial LEO/civilian ratio on the standard deviation might only manifest itself at larger numbers of iterations. Also the smaller number of runs per parameter value might affect the result. The result of the similar analysis done for the modified model is shown in figure ??. It

Table 5: Sensitivity indices computed for the modified model input parameters.

Parameter	S_i	S_i^T
1	0.3993	1.0651
2	-0.1700	0.7317
3	-0.2336	0.8763
4	-0.1632	0.7517
5	-0.1535	0.7772
6	-0.1971	0.7879
7	-0.1700	0.7482
8	-0.2087	0.8419
9	-0.2013	0.7784
10	-0.2031	0.8355
11	-0.0380	0.9466
12	-0.1666	0.8387
13	-0.1948	0.7765

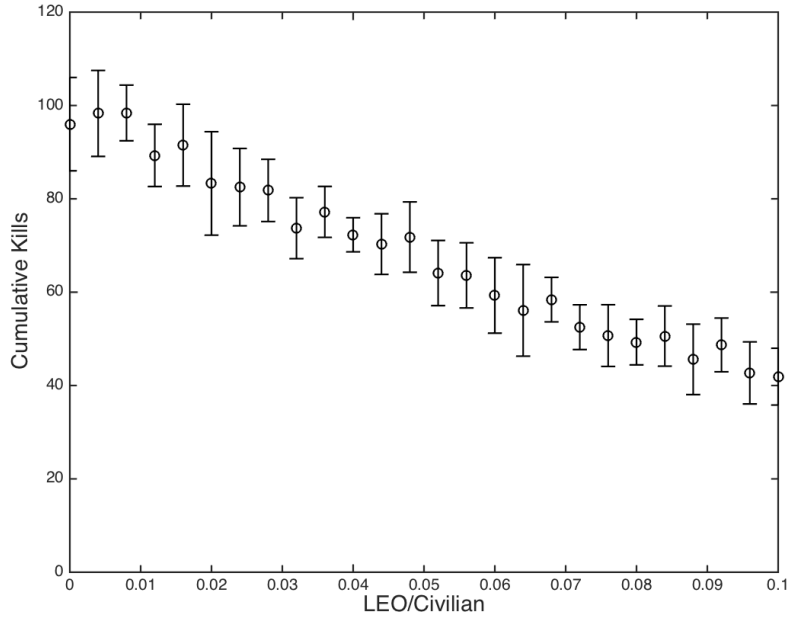


Figure 1: Effect of the initial LEO/civilian ratio on the cumulative number of kills at the end of the simulation using the base model. As expected, the mean formed out of 10 runs decreases with increasing LEO/civ ratio. The standard deviation on the other hand remains constant. It is shown via the error bars.

does not differ from that using the base model. This implies that the heterogeneity of the perceived legitimacy and the violence threshold as well as the updating of the legitimacy has no effect on the dependency of the cumulative number of kills on the initial LEO/civilian ratio, at least not at those values of all the other parameters.

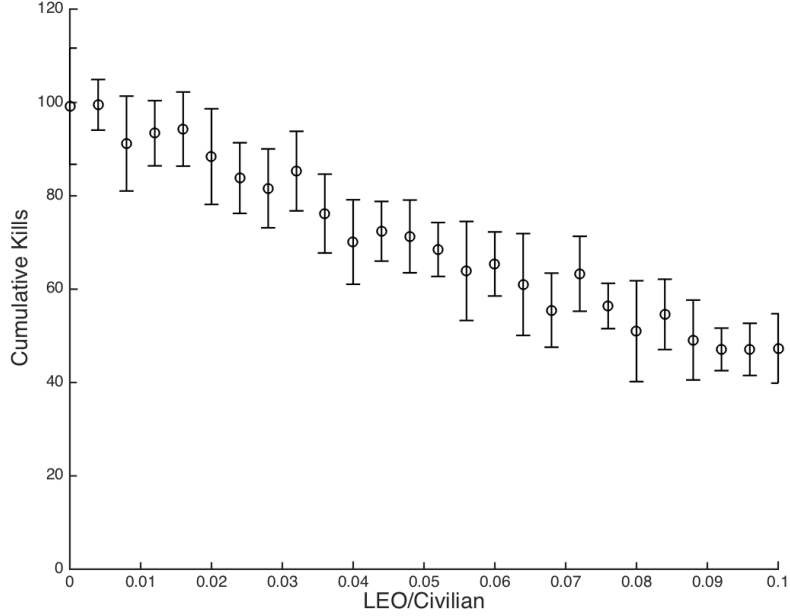


Figure 2: Effect of the initial LEO/civilian ratio on the cumulative number of kills at the end of the simulation using the modified model. Basically the same result as with the base model is obtained. Again, the mean formed out of 10 runs decreases with increasing LEO/civ ratio. The standard deviation shown through remains fairly constant.

6.2.2 Dependence on the Legitimacy and the Violence Threshold

7 Summary and Outlook

8 References

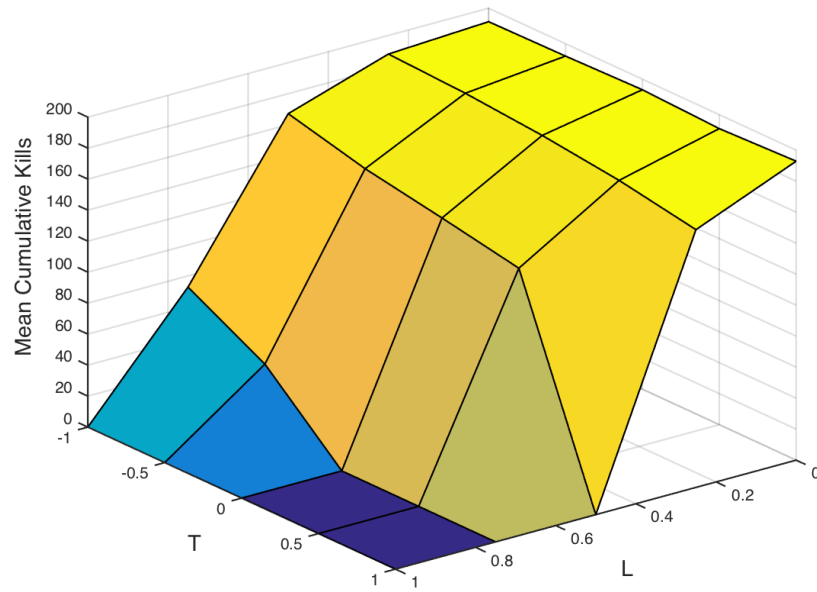


Figure 3: Dependence of the mean cumulative number of kills on the two parameters L and T of the base model. The parameter L seems to be more influential than T for the selected values of the other model parameters.

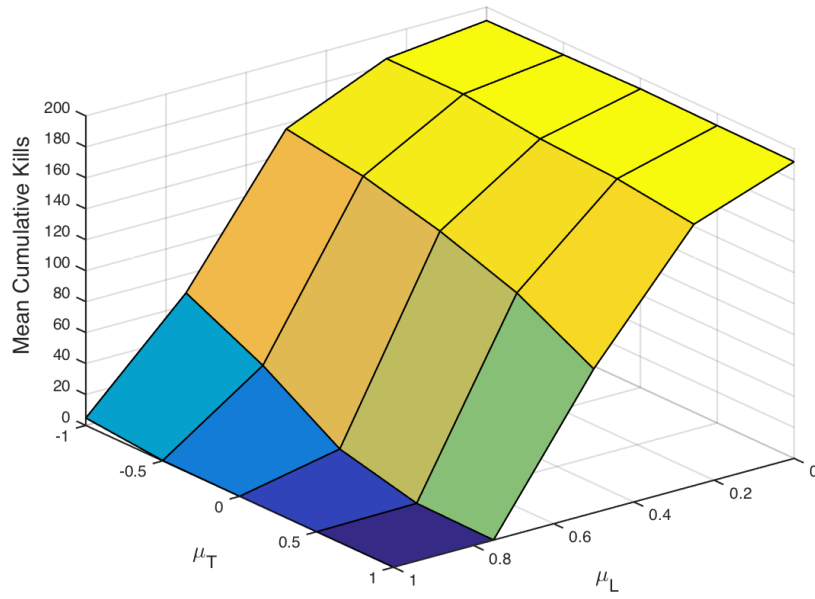


Figure 4: Dependence of the mean cumulative number of kills on the two parameters μ_L and μ_T of the modified model. As in the base model, the output shows more sensitivity towards the parameter L than the parameter T for the selected values of the other model parameters.